

Université de Sherbrooke

Les 12 isoformes de HNF4 α sont des régulateurs transcriptionnels distincts

Par
Élie Lambert
Programme de maîtrise en biologie cellulaire

Mémoire présenté à la Faculté de médecine et des sciences de la santé
en vue de l'obtention du grade de maître ès sciences (M.Sc.)
en biologie cellulaire

Sherbrooke, Québec, Canada
Janvier 2019

Membres du jury d'évaluation
Pr François-Michel Boisvert, Département d'anatomie et de biologie cellulaire, FMSS
Pr François Boudreau, Département d'anatomie et de biologie cellulaire, FMSS
Pr Claude Asselin, Département d'anatomie et de biologie cellulaire, FMSS
Pr Nicolas Gévry, Département de biologie, Faculté des sciences, UdeS

© Élie Lambert, 2019

À mes parents

Pour leur support inconditionnel

depuis le début

"And now that you don't have to be perfect, you can be good."

— John Steinbeck, East of Eden

RÉSUMÉ

Les 12 isoformes de HNF4 α sont des régulateurs transcriptionnels possédant des fonctions distinctes

Par
Élie Lambert
Programme de maîtrise en biologie cellulaire

Mémoire présenté à la Faculté de médecine et des sciences de la santé en vue de l'obtention du diplôme de maître ès sciences (M.Sc.) en biologie cellulaire, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec, Canada, J1H 5N4

HNF4 α est un récepteur nucléaire régulant la transcription de gènes impliqués principalement au niveau du développement, la différenciation cellulaire et le métabolisme. Des fonctions opposées pour les deux classes d'isoformes P1 et P2 de HNF4 α ont récemment été mises en lumière. Ces classes regroupent 12 variants de HNF4 α pouvant être exprimés par l'utilisation de deux promoteurs et par épissage alternatif. Jusqu'à maintenant, la caractérisation de ce facteur de transcription a fait abstraction de cette diversité et est demeurée confinée à l'étude d'une fraction des isoformes. Nous avons donc voulu éclaircir la situation en caractérisant spécifiquement les fonctions transcriptionnelles des 12 isoformes de HNF4 α . Nous avons généré à cet effet des lignées stables exprimant chaque isoforme de HNF4 α dans les cellules HCT 116.

Nous avons analysé l'ensemble du transcriptome associé à chaque isoforme par séquençage de l'ARN, ainsi que leur protéome par une approche BioID couplée à la spectrométrie de masse quantitative. Nous avons noté des différences majeures au niveau de la fonction transcriptionnelle des 12 isoformes. Les isoformes α 4, α 5 et α 6 ont été caractérisées pour la première fois, et montrent une capacité transcriptionnelle fortement réduite. Nous avons démontré que ces isoformes sont incapables de reconnaître l'élément de réponse consensus de HNF4 α . Les isoformes α 1 et α 2 sont les plus puissants régulateurs de l'expression génique, alors que l'isoforme α 3 présente une activité considérablement réduite. Plusieurs facteurs de transcription et corégulateurs ont été identifiés en tant que partenaires potentiels spécifiques pour certaines isoformes de HNF4 α . Le corépresseur IRF-2BP2 interagit spécifiquement avec les isoformes possédant la forme longue du domaine F de HNF4 α . Cette interaction spécifique pourrait expliquer la grande quantité de gènes modulés négativement par α 1 et α 2 comparativement à α 3.

L'analyse faisant l'intégration de la vaste quantité de données transcriptomiques et protéomiques générées au cours de ce projet permettra dans le futur d'identifier des mécanismes de régulation transcriptionnelle spécifiques à certaines isoformes de HNF4 α .

Mots clés : HNF4 α , isoformes, transcription, protéomique, récepteur nucléaire

TABLE DES MATIÈRES

Résumé.....	i
Table des matières	ii
Liste des figures	vii
Liste des tableaux	ix
Liste des abréviations	xi
INTRODUCTION	1
1. L'expression des gènes	1
1.1 Organisation du matériel génétique.....	1
1.2 La transcription	2
1.3 Facteurs de transcription.....	3
1.4 Corégulateurs transcriptionnels.....	4
1.4.1 Coactivateurs transcriptionnels.....	4
1.4.2 Corépresseurs transcriptionnels.....	5
2. Les récepteurs nucléaires	7
2.1 Structure et classification des récepteurs nucléaires.....	7
2.2 Activation et fonctions des récepteurs nucléaires.....	8
2.3 Isoformes des récepteurs nucléaires	10
3. Le récepteur nucléaire HNF4α	11
3.1 Structure et domaines de HNF4α	11
3.2 Mode d'action de HNF4α	15
3.3 Processus biologiques régulés par HNF4α	18
3.4 Régulation de HNF4α	20
4. Les isoformes de HNF4α.....	21
4.1 Expression différentielle des classes d'isoformes P1- et P2-HNF4α.....	21
4.2 Identification des isoformes de HNF4α.....	23
4.3 Activité transcriptionnelle des isoformes de HNF4α	26
5. Hypothèse et objectifs du projet de recherche	27

MATÉRIEL ET MÉTHODES.....	29
1. Clonage moléculaire.....	29
1.1 Clonage des isoformes de HNF4α.....	29
1.1.1 Clonage de la séquence commune dans pENTR11.....	29
1.1.2 Insertion des extrémités N-terminales dans pENTR11.....	30
1.1.3 Insertion des extrémités C-terminales dans pENTR11.....	31
1.1.4 Réaction LR dans les vecteurs d'expression pcDNA-DEST47, pgLAP5.2 et pgLAP5.2-BioID2-3Xmyc	31
1.2 Clonage du plasmide pgLAP5.2-BioID2-3Xmyc	32
1.3 Clonage du plasmide pgLAP5.2-BioID2-3Xmyc-vide	32
1.4 Clonage des gènes codant pour les protéines ayant un potentiel d'interaction avec les isoformes de HNF4α	33
2. Culture cellulaire.....	34
2.1 Lignées cellulaires HCT 116, HCT 116 Flp-In T-REx, HCT 116 HNF4α(1-12)- GFP, HCT 116 HNF4α(1-12)-BioID2-3Xmyc et HCT 116 BioID2-3Xmyc-vide.....	34
2.2 Lignée cellulaire 293T	35
2.3 Lignées cellulaires Caco-2/15, Capan-2, et Hep G2	35
2.4 Lignées cellulaires AsPC-1, COLO 205 et DLD-1	35
2.5 Lignée cellulaire HT-29.....	36
2.6 Lignée cellulaire LoVo	36
2.7 Lignée cellulaire T84.....	36
2.8 Culture cellulaire en milieu SILAC.....	37
3. Transfections	38
3.1 Transfections transitoires	38
3.2 Génération de lignées cellulaires stables	38
4. Immunofluorescence	39
5. Extraction et analyse des ARN	40
5.1 Extraction des ARN totaux à partir de cellules en culture.....	40
5.1.1 Extraction phénol-chloroforme	41
5.1.2 Extraction sur colonne à membrane de silice.....	41
5.2 Transcription inverse.....	42
5.3 PCR (<i>Polymerase Chain Reaction</i>) semi-quantitative	43
5.3.1 Expression des isoformes dans différentes lignées cellulaires cancéreuses humaines	43
5.3.2 Expression des isoformes dans différents tissus humains du système digestif	43

5.4 PCR quantitative en temps réel	44
6. Extraction protéique	45
6.1 Extraction des protéines totales	45
6.2 Extraction des protéines nucléaires	45
7. Immunobuvardage (Western blot)	46
7.1 Électrophorèse des extraits totaux de protéines sur gel de polyacrylamide SDS	46
7.2 Transfert des protéines	46
7.3 Détection des protéines.....	47
8. Retard sur gel	48
8.1 Hybridation des sondes biotinylées.....	48
8.2 Réactions de liaison	49
8.3 Migration sur gel de polyacrylamide non dénaturant	49
8.4 Transfert sur membrane de nylon chargée positivement	50
8.5 Détection des sondes biotinylées.....	50
9. Analyse du transcriptome des lignées HCT 116 HNF4α(1-12)-GFP	51
9.1 Préparation des ARN totaux.....	51
9.2 Préparation des bibliothèques et séquençage des ARN.....	51
9.3 Alignement des résultats de séquençage	52
10. Spectrométrie de masse.....	52
10.1 Culture cellulaire et induction des différentes constructions	52
10.2 <i>Pull-down</i> des protéines biotinylées.....	53
10.3 Réduction, alkylation et digestion des protéines.....	55
10.4 Purification des peptides sur colonne C18	55
10.5 Analyse LC-MS/MS des échantillons peptidiques	56
10.6 Identification des protéines par analyse MaxQuant	57
11. Validation des interactions par co-immunoprécipitation	57
11.1 Co-transfection de HNF4α(1-12)-GFP et des gènes partenaires MS-FLAG ...	57
11.2 Co-immunoprécipitation	58
12. Production et infection lentivirales	59
12.1 Production lentivirale.....	59
12.2 Infection lentivirale	60
RÉSULTATS	61

1. Générer et valider des lignées stables exprimant de façon spécifique chacune des 12 isoformes de HNF4a	61
1.1 Expression endogène des isoformes	61
1.1.1 Les isoformes sont exprimées différemment dans les tissus gastro-intestinaux humains	61
1.1.2 L'expression des isoformes est hautement variable dans plusieurs lignées cellulaires cancéreuses humaines.....	63
1.2 Validation des lignées HCT 116 HNF4a(1-12)-GFP et HCT 116 HNF4a(1-12)-BioID2-3Xmyc.....	65
1.2.1 Niveaux d'expression protéique des isoformes de HNF4a en fusion avec les étiquettes protéiques GFP et BioID2-3Xmyc	65
1.2.2 Les constructions impliquant les différentes isoformes de HNF4a se localisent exclusivement dans le noyau des cellules HCT 116	67
1.2.3 Les isoformes de HNF4a n'ont pas toutes la même capacité de liaison à l'ADN	69
1.2.4 HNF4a active la transcription de ses gènes cibles de façon variable en fonction de ses isoformes	71
2. Analyser l'impact fonctionnel de l'expression des isoformes sur le transcriptome dans le contexte des cellules HCT 116.....	73
2.1 Les isoformes qui partagent le même domaine A/B sont associées à des profils transcriptomiques plus similaires.....	73
2.2 Comparaison du transcriptome des cellules HCT 116 exprimant chacune des 12 isoformes de HNF4a	75
3. Étudier les partenaires d'interaction spécifiques pour chaque isoforme.....	79
3.1 Les 12 isoformes de HNF4a possèdent des réseaux d'interactions distincts, formés en majeure partie de facteurs de transcription et de corégulateurs transcriptionnels	79
3.2 Identification de partenaires potentiels spécifiques à certaines isoformes de HNF4a.....	84
3.3 Validation des interactions (Co-IP)	86
4. Identifier des corégulateurs spécifiques à certaines isoformes pour expliquer les variations dans leurs fonctions transcriptionnelles	89
4.1 Le corépresseur IRF-2BP2 interagit préférentiellement avec les isoformes de HNF4a possédant la forme longue du domaine F	89
DISCUSSION.....	92
1. Le profil d'expression de HNF4a est hautement variable entre ses 12 isoformes	92
1.1 Distribution tissulaire des isoformes	92
1.2 Expression protéique des isoformes.....	93

2. Les 12 isoformes de HNF4α modulent de façon distincte le transcriptome des cellules HCT 116	96
3. La spectrométrie de masse quantitative révèle des réseaux d'interactions spécifiques aux isoformes de HNF4α	98
3.1 Les fonctions différentielles des isoformes sont finement régulées via leur interaction avec un sous-ensemble de régulateurs transcriptionnels.....	98
3.2 L'approche BioID permet l'identification de plusieurs protéines ayant une réelle interaction avec HNF4 α	99
4. Le corépresseur IRF-2BP2 pourrait réguler spécifiquement les isoformes ne possédant pas la forme courte du domaine F.....	100
5. Plusieurs nouveaux interactants potentiels de HNF4α ont été identifiés de façon indépendante des isoformes exprimées	104
6. La fonction des isoformes α4-5-6 est indépendante de leur liaison au site consensus DR1 de HNF4α	105
CONCLUSION GÉNÉRALE.....	108
REMERCIEMENTS	110
RÉFÉRENCES.....	112
ANNEXE A - Nomenclature des récepteurs nucléaires.....	125
ANNEXE B - La séquence réelle de HNF4α4	126
ANNEXE C - Annotations des isoformes de HNF4α à travers différentes bases de données	130
ANNEXE D - Séquences des plasmides HNF4α.....	131
ANNEXE E - Conditions PCR	137
ANNEXE F - Lectures de séquençage de l'ARN.....	145
ANNEXE G - Enrichissement SILAC des sous-unités des complexes BAF/PBAF	146

LISTE DES FIGURES

INTRODUCTION

Figure 1 : Les domaines fonctionnels du récepteur nucléaire HNF4α	12
Figure 2 : HNF4α lie l'ADN sous forme d'homodimère.....	13
Figure 3 : Les coactivateurs et corépresseurs interagissant avec HNF4α régulent son activité transcriptionnelle	17
Figure 4 : Les classes d'isoformes P1- et P2-HNF4α varient uniquement au niveau de leur domaine A/B	23
Figure 5 : Le gène <i>HNF4A</i> peut mener à l'expression de 12 isoformes, par l'utilisation de promoteurs alternatifs et par épissage alternatif	25

RÉSULTATS

Figure 6 : Les isoformes de HNF4α sont exprimées à des niveaux variables dans différents tissus du tractus gastro-intestinal humain	62
Figure 7 : Les isoformes de HNF4α sont exprimées à des niveaux variables dans différentes lignées cellulaires cancéreuses humaines	64
Figure 8 : Les constructions des 12 isoformes de HNF4α en fusion avec les étiquettes protéiques GFP et BioID2-3Xmyc sont exprimées suite à l'induction des lignées stables dans les cellules HCT 116	66
Figure 9 : Les lignées stables HCT 116 exprimant les 12 isoformes de HNF4α en fusion avec les étiquettes protéiques GFP et BioID2-3Xmyc montrent que celles-ci sont localisées dans le compartiment nucléaire.....	68
Figure 10 : Les isoformes de HNF4α ne peuvent pas toutes reconnaître la séquence consensus DR1 du récepteur nucléaire	70
Figure 11 : Des gènes cibles connus de HNF4α sont transactivés à différents niveaux en fonction des isoformes exprimées.....	72
Figure 12 : L'analyse de la variance entre les ensembles de données révèle des profils transcriptomiques plus similaires pour les isoformes ayant un domaine A/B commun.....	74
Figure 13 : L'expression des 12 isoformes de HNF4α provoque des modulations hautement variables du transcriptome des cellules HCT 116.....	76

Figure 14 : Représentation « volcano plot » de l'expression différentielle des gènes	78
Figure 15 : Stratégie d'identification des partenaires protéiques des isoformes de HNF4a par l'approche BioID couplée à la spectrométrie de masse quantitative.....	80
Figure 16 : Le protéome est particulièrement similaire chez les isoformes partageant le même domaine A/B	82
Figure 17 : Les interactants potentiels pour les 12 isoformes sont principalement des corégulateurs transcriptionnels et des facteurs de transcription.....	83
Figure 18 : Plusieurs protéines interagissent spécifiquement avec les isoformes qui possèdent un domaine A/B ou F commun	85
Figure 19 : Listes de protéines sélectionnées et validées pour leur interaction avec les isoformes de HNF4a.....	87
Figure 20 : Validation des interactions entre les isoformes de HNF4a et cinq protéines identifiées par l'approche BioID.....	88
Figure 21 : Les corépresseurs IRF-2BP1 et IRF-2BP2 sont enrichis spécifiquement par les isoformes qui ne possèdent pas la forme courte du domaine F	90
Figure 22 : Validation de l'inhibition du corépresseur IRF-2BP2 via interférence par ARN..	91

ANNEXES

Figure 23 : Alignement de la séquence de l'isoforme HNF4a4 publiée initialement avec la séquence de l'isoforme HNF4a2	126
Figure 24 : Reconstitution de la séquence de HNF4a4 à partir du gène humain <i>HNF4A</i>	127
Figure 25 : L'utilisation du codon initiateur alternatif pour l'isoforme $\alpha 2$ produit une protéine tronquée	128
Figure 26 : La séquence réelle de l'isoforme HNF4a4	129

LISTE DES TABLEAUX

INTRODUCTION

Tableau 1 : Les régions activatrices AF-1 et AF-2 coordonnent les interactions de HNF4α avec divers corégulateurs transcriptionnels 14

Tableau 2 : L'expression des classes d'isoformes P1- et P2-HNF4α est compartimentée selon les tissus 23

MATÉRIEL ET MÉTHODES

Tableau 3 : Liste des anticorps utilisés pour les essais d'immunofluorescence..... 40

Tableau 4 : Liste des réactifs utilisés dans le mélange réactionnel de transcription inverse 42

Tableau 5 : Liste des anticorps utilisés pour la détection des protéines par immunobuvardage 48

Tableau 6 : Liste des oligonucléotides utilisés pour le retard sur gel 49

Tableau 7 : Combinaisons des lignées cellulaires en conditions SILAC pour le *pull-down* des protéines biotinylées 54

Tableau 8 : Vecteurs d'expression des shARN et leur séquence cible 60

ANNEXES

Tableau 9 : Nomenclature classique et uniformisée des récepteurs nucléaires, ainsi que leurs ligands identifiés jusqu'à présent.....125

Tableau 10 : Identification des isoformes de HNF4α dans les bases de données RefSeq, Uniprot et Ensembl130

Tableau 11 : Oligonucléotides utilisés pour l'amplification de la séquence commune de HNF4α137

Tableau 12 : Réactifs utilisés pour l'amplification de la séquence commune de HNF4α137

Tableau 13 : Paramètres PCR d'amplification de la séquence commune de HNF4α137

Tableau 14 : Oligonucléotides utilisés pour l'amplification du plasmide pgLAP5.2.....	138
Tableau 15 : Oligonucléotides utilisés pour amplifier le vecteur pgLAP5.2-BioID2-3Xmyc	138
Tableau 16 : Paramètres PCR d'amplification du vecteur pgLAP5.2-BioID2-3Xmyc-vide	138
Tableau 17 : Liste des oligonucléotides utilisés pour l'amplification des gènes partenaires MS	139
Tableau 18 : Réactifs utilisés pour l'amplification des gènes partenaires MS.....	140
Tableau 19 : Paramètres d'amplification des gènes partenaires MS	140
Tableau 20 : Liste des oligonucléotides utilisés pour l'amplification des 12 isoformes de HNF4a et des gènes de référence	141
Tableau 21 : Réactifs utilisés pour l'amplification des isoformes de HNF4a par PCR	142
Tableau 22 : Paramètres PCR d'amplification des isoformes de HNF4a	142
Tableau 23 : Liste des oligonucléotides utilisés pour l'amplification par RT-qPCR	143
Tableau 24 : Réactifs utilisés pour l'amplification par RT-qPCR.....	143
Tableau 25 : Conditions d'amplification par RT-qPCR	144
Tableau 26 : Nombre de lectures obtenues par séquençage Novaseq pour chaque échantillon exprimant ou non les isoformes de HNF4a en triplicatas	145
Tableau 27 : Enrichissement des membres des complexes de remodelage de la chromatine SWI/SNF BAF/PBAF par les 12 isoformes de HNF4a	146

LISTE DES ABRÉVIATIONS

ADN	Acide désoxyribonucléique
AF-1/2.....	Fonction d'activation 1/2 / <i>Activation function 1/2</i>
AMPK	<i>5'-AMP-activated protein kinase</i>
AP-1.....	<i>Activator protein 1</i>
AR	Récepteur des androgènes / <i>Androgen receptor</i>
ARN	Acide ribonucléique
ARNm	ARN messenger
ATAC	<i>Ada2a-containing HAT complex</i>
ATF-2.....	<i>Activating transcription factor 2</i>
BAF.....	<i>BRG1/BRM-associated factor</i>
BioID	Identification de la biotinylation dépendante de la proximité
CAR	Récepteur constitutif des androstanes / <i>Constitutive androstane receptor</i>
CARM1	<i>Coactivator-associated arginine methyltransferase 1</i>
CBP.....	<i>CREB-binding protein</i>
CDYL.....	<i>Chromodomain Y like</i>
CHD	<i>Chromodomain helicase DNA-binding</i>
Co-IP	Co-immunoprécipitation
COUP-TFI	<i>Chicken ovalbumin upstream promoter transcription factor I</i>
COUP-TFII	<i>Chicken ovalbumin upstream promoter transcription factor II</i>
DBD	Domaine de liaison à l'ADN / <i>DNA-binding domain</i>
DNMT1.....	<i>DNA methyltransferase 1</i>
DNMT3.....	<i>DNA methyltransferase 3</i>
DR	Répétition directe / <i>Direct repeat</i>
ER α	Récepteur alpha des œstrogènes / <i>Estrogen receptor alpha</i>
FOXA1.....	<i>Forkhead box A1</i>
FOXA2.....	<i>Forkhead box A2</i>
FRA-1.....	<i>Fos-related antigen 1</i>
FT.....	Facteur de transcription
GATA6	<i>GATA binding protein 6</i>
GO-BP	<i>Gene ontology - Biological process</i>
GO-MF	<i>Gene ontology - Molecular function</i>
GR	Récepteur des glucocorticoïdes / <i>Glucocorticoid receptor</i>
GTF	Facteur de transcription général / <i>General transcription factor</i>
HAT	Histone acétyltransférase
HDAC	Histone désacétylase
HNF	<i>Hepatocyte nuclear factor</i>
HNF1 α	<i>Hepatocyte nuclear factor 1 alpha</i>
HNF1 β	<i>Hepatocyte nuclear factor 1 beta</i>
HNF4 α	<i>Hepatocyte nuclear factor 4 alpha</i>
HNF4 γ	<i>Hepatocyte nuclear factor 4 gamma</i>
HNF6	<i>Hepatocyte nuclear factor 6</i>
HRE	Élément de réponse aux hormones / <i>Hormone response element</i>
INO80	<i>Inositol-requiring protein 80</i>
ISWI	<i>Imitation SWI</i>
KAT2B.....	<i>Lysine acetyltransferase 2B</i>

LBD.....	Domaine de liaison au ligand / <i>Ligand-binding domain</i>
MBD3.....	<i>Methyl-CpG-binding domain protein 3</i>
MeCP2	<i>Methyl-CpG-binding protein 2</i>
MLL.....	<i>Myeloid/lymphoid or mixed-lineage leukemia</i>
MPT	Modification post-traductionnelle
NCoA-1	<i>Nuclear receptor coactivator 1</i>
NCoA-2	<i>Nuclear receptor coactivator 2</i>
NCoA-3	<i>Nuclear receptor coactivator 3</i>
NCoR1.....	<i>Nuclear receptor co-repressor 1</i>
NCoR2.....	<i>Nuclear receptor co-repressor 2</i>
NFAT.....	<i>Nuclear factor of activated T-cells</i>
NF- κ B.....	<i>Nuclear factor kappa-light-chain-enhancer of activated B cells</i>
NLS.....	Signal de localisation nucléaire / <i>Nuclear localization signal</i>
NuA4.....	<i>Nucleosome acetyltransferase of H4</i>
NuRD	<i>Nucleosome Remodeling Deacetylase</i>
OC2	<i>One cut domain family member 2</i>
PBAF.....	<i>Polybromo-associated BAF</i>
PCA.....	<i>Principal component analysis</i>
PcG.....	<i>Polycomb-group</i>
PDX1.....	<i>Pancreas/duodenum homeobox protein 1</i>
PIC	Complexe de pré-initiation / <i>Preinitiation complex</i>
PKA.....	Protéine kinase A
PKC.....	Protéine kinase C
PolII.....	ARN polymérase II
PR	Récepteur de la progestérone / <i>Progesterone receptor</i>
PRMT1	<i>Protein arginine N-methyltransferase 1</i>
RAR	Récepteur de l'acide rétinoïque / <i>Retinoic acid receptor</i>
RN	Récepteur nucléaire
ROR α	Récepteur alpha apparenté au RAR / <i>RAR-related orphan receptor alpha</i>
RXR	Récepteur X des rétinoïdes / <i>Retinoid X receptor</i>
SILAC.....	<i>Stable isotope labeling with amino acids in cell culture</i>
SNF2.....	<i>Sucrose non-fermentable 2</i>
SREBP-1	<i>Sterol regulatory element-binding protein 1</i>
SWI/SNF	<i>Switch/Sucrose non-fermentable</i>
TAF.....	<i>TATA-binding protein associated factor</i>
TBP.....	<i>TATA-binding protein</i>
TCF4.....	<i>Transcription factor 4</i>
TNF α	<i>Tumour necrosis factor alpha</i>
TPM	Transcrits par million
TR	Récepteur des hormones thyroïdiennes / <i>Thyroid hormone receptor</i>

INTRODUCTION

1. L'expression des gènes

Le génome constitue l'ensemble du matériel génétique d'un organisme, encodé sous forme d'ADN (acide désoxyribonucléique). L'expression génique implique une première étape de transcription, consistant en la production d'un brin d'ARN (acide ribonucléique) copié à partir de la séquence d'un gène. Le matériel génétique est constitué de séquences codantes, qui mènent à la transcription de brins d'ARN messager (ARNm) traduits par la suite en protéines, ainsi que de séquences non codantes. Ces séquences regroupent toutes les régions qui ne sont pas traduites en protéines, certaines d'entre elles pouvant toutefois être transcrites en molécules fonctionnelles d'ARN (ARN de transfert, ARN ribosomiaux, micro ARN, petits ARN nucléaires, etc.). L'ensemble de l'information génétique est présente de façon identique dans tous les types cellulaires d'un organisme. Différents mécanismes de régulation transcriptionnelle mènent à l'expression spécifique d'un ensemble de gènes, permettant à une cellule de se spécialiser et d'acquérir une identité qui lui est propre (Alberts *et al.*, 2015).

1.1 Organisation du matériel génétique

L'ADN est contenu dans le noyau de la cellule, et est associé à différentes protéines histones et non-histones sous forme de chromatine. L'interaction des histones et d'autres protéines avec l'ADN permet de réguler son niveau de compaction. Les nucléosomes, qui représentent l'unité minimale de la chromatine, sont formés d'un octamère entre deux dimères d'histones H2A-H2B et d'un tétramère (H3-H4)₂ autour duquel sont entourées 146 paires de bases d'ADN (Luger *et al.*, 1997). Au moment de la mitose, la chromatine atteint une condensation maximale, prenant alors la forme de chromosomes, afin de faciliter la propagation du matériel génétique lors de la division cellulaire. De façon globale, la chromatine peut être distinguée selon son degré de compaction en deux types, soit l'euchromatine et l'hétérochromatine. Sous forme d'euchromatine, l'ADN est dans un état relâché et est donc associé à une transcription active, alors que l'hétérochromatine est

associée à une répression transcriptionnelle due au niveau plus important de compaction de l'ADN (Woodcock et Ghosh, 2010).

1.2 La transcription

Le processus de transcription est un point de contrôle crucial de l'expression des gènes puisqu'il en constitue la première étape. Dans le cas des séquences codantes, ce processus se termine par la synthèse d'une molécule d'ARNm, qui sera éventuellement traduite en une protéine. On peut diviser le processus de transcription en trois phases, soit l'initiation, l'élongation et la terminaison. L'initiation de la transcription nécessite le recrutement de l'ARN polymérase de type II (PolII) au promoteur des gènes sous forme d'un complexe de pré-initiation (PIC - *Preinitiation complex*) (Krasnov *et al.*, 2016). Cette liaison au promoteur implique que cette région doit être accessible aux diverses protéines membres du PIC, et donc être sous forme d'euchromatine. Le PIC est composé minimalement de la PolII et des facteurs de transcription généraux (GTF - *General transcription factors*), auxquels peuvent s'ajouter d'autres protéines tel le complexe Mediator (Maston *et al.*, 2006 ; Poss *et al.*, 2013).

L'assemblage du PIC est séquentiel, et débute par la liaison au promoteur du GTF TFIID, un complexe formé du GTF TBP (*TATA-binding protein*) et de plusieurs sous-unités TAF (*TATA-binding protein associated factors*). Cette liaison de TFIID au promoteur est par la suite stabilisée par TFIIA et TFIIB. La liaison subséquente de la PolII et TFIIF, puis de TFIIIE et TFIIH, complète la formation du PIC (Sainsbury *et al.*, 2015). Via l'activité ATPase de TFIIH, l'ADN double-brin est alors ouvert et forme une « bulle de transcription » d'ADN simple-brin, permettant l'initiation de la transcription par la PolII. L'élongation est rapidement et transitoirement arrêtée, puisque la PolII doit d'abord surmonter la clairance du promoteur. Celle-ci se définit comme une phase initiale de l'élongation où le complexe transcriptionnel est instable et tend à relâcher l'ARNm en élongation. Les facteurs TFIIH et TFIIF facilitent la transition de la PolII vers une phase d'élongation beaucoup plus rapide et efficace. L'étape finale de la transcription consiste en sa terminaison, lorsque l'ARNm est clivé, polyadénylé et transporté au cytoplasme pour y être traduit en protéine (Sims *et al.*, 2004).

Bien que le recrutement du PIC se fasse au niveau du promoteur des gènes devant être transcrits, d'autres régions appelées amplificateurs (ou *enhancers*) peuvent réguler de manière positive ou négative la transcription. Ces régions sont généralement distales, pouvant être situées à plusieurs centaines de milliers de paires de bases en amont ou en aval d'un promoteur, mais peuvent en être rapprochées par une courbure de la structure chromatinienne. Les amplificateurs peuvent alors interagir avec la machinerie de transcription générale et ainsi participer à la régulation du niveau de transcription de leurs gènes cibles (Ong et Corces, 2011).

1.3 Facteurs de transcription

Approximativement 1 400 gènes, correspondant à 6 % des gènes encodant des protéines, codent pour des facteurs de transcription (FT). Ces protéines peuvent lier de façon spécifique certaines séquences d'ADN et réguler positivement ou négativement le recrutement de la machinerie d'initiation de la transcription aux promoteurs de leurs gènes cibles (Vaquerizas *et al.*, 2009). Les facteurs de transcription sont communément regroupés en fonction de la structure de leur domaine de liaison à l'ADN (DBD - *DNA binding domain*), tels les doigts de zinc, les homéodomains et les motifs en hélice-boucle-hélice (Luscombe *et al.*, 2000). Via la reconnaissance de séquences spécifiques par leur DBD, les FT peuvent ainsi dicter l'expression génique des cellules dans lesquelles ils sont eux-mêmes exprimés. Les FT peuvent coopérer entre eux pour lier l'ADN, par des liaisons protéine-protéine et via la proximité de leurs éléments de réponse, permettant ainsi une meilleure spécificité de liaison. Cette notion de coopérativité est aussi importante pour expliquer la liaison des FT au niveau de la chromatine plus compactée. En effet, l'inaccessibilité de l'hétérochromatine peut être neutralisée par l'action combinatoire de plusieurs FT se liant à leurs sites de liaison de façon simultanée (Reiter *et al.*, 2017). Cependant, la liaison des FT à leurs éléments de réponse nécessite généralement une interaction avec l'environnement de la chromatine entourant ces séquences régulatrices (Lonard et O'Malley, 2012). Diverses protéines, appelées corégulateurs transcriptionnels, médient cette interaction de plusieurs manières, tels le remodelage de la chromatine et la modification covalente des histones. Les corégulateurs, distingués en coactivateurs et corépresseurs, participent ainsi à la régulation de l'expression des gènes en tant qu'intermédiaires des interactions entre les FT, leur environnement chromatinien et la machinerie de transcription générale. Certains FT peuvent

également recruter directement le complexe de transcription basale, alors que d'autres bloquent l'action d'autres FT par liaison compétitive à un élément de réponse commun (Lambert *et al.*, 2018).

1.4 Corégulateurs transcriptionnels

La fonction de régulation de l'expression génique par les facteurs de transcription est modulée par plusieurs corégulateurs transcriptionnels. Jusqu'à maintenant, plus de 400 protéines sont considérées comme des corégulateurs de la transcription, et peuvent être catégorisées en tant que coactivateurs et corépresseurs (Dasgupta *et al.*, 2014). Ces protéines ne peuvent lier l'ADN directement, et doivent donc interagir avec les FT afin d'exercer leurs fonctions régulatrices. La plupart des corégulateurs ne peuvent fonctionner de manière individuelle, et doivent s'associer afin de former de larges complexes multiprotéiques, favorisant le rapprochement de différents mécanismes de régulation utilisés par ces protéines. Certains membres de ces complexes peuvent interagir directement avec différents facteurs de transcription, menant au recrutement du complexe corégulateur à certaines régions génomiques spécifiquement régulées par ces FT. D'autres corégulateurs possèdent différentes activités enzymatiques menant à une modification covalente des histones ou au remodelage de la chromatine. Un troisième groupe de corégulateurs permet la présentation de substrats nécessaires à l'activité enzymatique du complexe, ainsi que la reconnaissance de modifications de la chromatine. Finalement, certains corégulateurs sont simplement considérés comme des adaptateurs participant à l'assemblage global de ces différents complexes multifonctionnels (Millard *et al.*, 2013).

1.4.1 Coactivateurs transcriptionnels

Les coactivateurs sont des protéines capables de lier des activateurs (facteurs de transcription) et de mener de façon directe ou indirecte à l'activation de l'expression des gènes. Certains coactivateurs, comme le complexe Mediator, peuvent recruter directement la machinerie de transcription générale. Ce complexe, formé d'une trentaine de sous-unités polypeptidiques, sert ainsi à faire le pont entre les facteurs de transcription, la PolII et les facteurs de transcription généraux, favorisant donc la formation du PIC (Yin et Wang, 2014). Un deuxième type de coactivateurs possède une activité enzymatique intrinsèque leur permettant de provoquer un relâchement dans la structure de la chromatine par la

modification post-traductionnelle de la queue des histones formant les nucléosomes (H2A, H2B, H3 et H4). L'acétylation des histones est une marque fortement associée à une transcription active des gènes, puisqu'elle mène à un état plus relâché de la chromatine. Plusieurs coactivateurs exhibent une activité histone acétyltransférase (HAT), telles les protéines CBP (*CREB-binding protein*), p300 et la famille p160 de coactivateurs, qui contient les protéines NCoA-1/2/3 (*Nuclear receptor coactivator 1/2/3*) (Näär *et al.*, 2001). L'état de méthylation des histones peut aussi affecter l'expression des gènes, et est régulé par différents coactivateurs, incluant les méthyltransférases CARM1 (*Coactivator-associated arginine methyltransferase 1*) et PRMT1 (*Protein arginine N-methyltransferase 1*), ainsi que les méthyltransférases de la famille MLL (*Myeloid/lymphoid or mixed-lineage leukemia*) (Hsia *et al.*, 2010 ; Gu et Lee, 2013). Une troisième classe de coactivateurs regroupe les complexes de remodelage de la chromatine dépendants de l'ATP. Ces complexes possèdent tous une sous-unité à activité ATPase de la famille SNF2 (*Sucrose non-fermentable 2*) des ADN hélicases. On recense quatre sous-familles distinctes, soit les complexes SWI/SNF (*Switch/Sucrose non-fermentable*), ISWI (*Imitation SWI*), CHD (*Chromodomain helicase DNA-binding*) et INO80 (*Inositol requiring*) (Hota et Bruneau, 2016). Ces complexes sont capables de déplacer les nucléosomes le long de l'ADN, de retirer des nucléosomes de la chromatine ou encore de changer leur composition en variants d'histones (Pollard *et al.*, 2016). Le recrutement par les FT de cette classe de coactivateurs vise donc à rendre accessible la séquence promotrice des gènes ciblés par les FT et ainsi permettre la formation du PIC au promoteur. Un autre modèle propose l'existence de complexes préformés entre les FT et ces coactivateurs, ce qui faciliterait le recrutement initial des FT à leurs éléments de réponse (Krasnov *et al.*, 2016).

1.4.2 Corépresseurs transcriptionnels

Les corépresseurs sont des protéines pouvant interagir non seulement avec les FT, mais également avec la chromatine, et menant à la répression des gènes. De manière analogue aux coactivateurs, les corépresseurs s'associent généralement pour former de larges complexes multiprotéiques. Ces complexes coordonnent des mécanismes communs de répression transcriptionnelle, impliquant la séquestration de GTF ainsi que la modification de la structure de la chromatine (Urnov *et al.*, 2001). Différentes enzymes s'associent ou font partie de ces complexes corépresseurs, telles les histones désacétylases (HDAC), les

histones H3K4 déméthylases, les histones H3K9 et H3K27 méthyltransférases ainsi que divers complexes de remodelage de la chromatine dépendants de l'ATP. Par opposition aux coactivateurs, ces enzymes répriment la transcription puisqu'elles mènent à un état plus compacté de la chromatine, ce qui empêche la liaison des FT à leurs éléments de réponse et l'assemblage du PIC au promoteur. Certaines ADN méthyltransférases comme DNMT1/3 (*DNA methyltransferase 1/3*) sont recrutées par les FT et causent l'hyperméthylation de leurs gènes cibles, et sont considérées comme un autre type de corépresseurs transcriptionnels (Stewart et Wong, 2009).

L'association des corépresseurs avec la chromatine représente un état stable, qui doit être contrebalancé par la liaison d'activateurs et de coactivateurs afin de permettre l'expression des gènes. Comme la répression des gènes implique un état compacté de la chromatine, comment font les différents complexes corépresseurs pour y accéder et y exercer leur fonction? Comme mentionné dans la section 1.3, certains FT peuvent lier l'hétérochromatine et de façon subséquente y recruter les corépresseurs. Cela demeure une explication partielle à cette question, et d'autres hypothèses ont été avancées pour y répondre. Au moment de la réplication, la structure immature de la chromatine pourrait permettre le recrutement et l'incorporation de corépresseurs à ce moment particulier. Également, certains complexes corépresseurs tel NuRD (*Nucleosome Remodeling Deacetylase*) possèdent une activité de remodelage de la chromatine dépendante de l'ATP, ainsi qu'une activité histone désacétylase, menant à la déstabilisation des nucléosomes. L'association des complexes corépresseurs avec la chromatine est médiée par leur liaison à certaines protéines composant les nucléosomes. Par exemple, les complexes corépresseurs SIN3-HDAC et NuRD s'associent respectivement avec les protéines liant les îlots CpG méthylés MeCP2 (*Methyl-CpG-binding protein 2*) et MBD3 (*Methyl-CpG-binding domain protein 3*). Le complexe PcG (*Polycomb-group*) interagit directement avec les histones formant l'octamère du nucléosome (Urnov *et al.*, 2001). Il peut notamment causer la tri-méthylation de la lysine 27 de l'histone H3, une marque associée à une répression transcriptionnelle (Margueron et Reinberg, 2001 ; Barski *et al.*, 2007). Deux autres types de corépresseurs caractérisés particulièrement pour leur interaction avec les FT de la superfamille des récepteurs nucléaires, NCoR1 (*Nuclear receptor co-repressor 1*) et NCoR2 (*Nuclear receptor co-repressor 2*), sont décrits plus en détail dans la section 2.2 de ce chapitre.

2. Les récepteurs nucléaires

2.1 Structure et classification des récepteurs nucléaires

Les récepteurs nucléaires (RN) représentent une classe de facteurs de transcription regroupant chez l'humain près d'une cinquantaine de protéines évolutivement reliées. La nomenclature entourant la superfamille des RN, basée sur leur phylogénie, fait état de 6 sous-familles composées de plusieurs groupes (Nuclear Receptors Nomenclature Committee, 1999) (se référer au tableau 9, annexe A, pour la liste complète). Les RN partagent une organisation structurale regroupant de cinq à six régions distinctes désignées de A à F (Robinson-Rechavi *et al.*, 2003). À ce jour, il demeure impossible d'élucider la structure tridimensionnelle complète de ces récepteurs, dû à la difficulté de purifier et de cristalliser ceux-ci. La conformation 3D de certains domaines, tels le domaine de liaison à l'ADN et le domaine de liaison au ligand, a cependant pu être établie. D'autres régions, comme le domaine A/B, sont plus difficiles à cristalliser dû à leur structure mal définie en matière de repliement (Wärnmark *et al.*, 2003).

La région A/B, située à l'extrémité N-terminale, est très peu conservée entre les RN, présentant une grande variabilité en ce qui a trait à leur longueur et leur séquence. On y retrouve une région de transactivation AF-1 (*Activation function*) qui est active indépendamment de la liaison au ligand. Cette région est connue pour permettre l'interaction du récepteur avec différents corégulateurs et autres facteurs de transcription. La région AF-1 de certains récepteurs stéroïdiens peut interagir avec une seconde région activatrice appelée AF-2, modulant ainsi leur interaction avec différents corégulateurs (Huang *et al.*, 2010).

Le domaine C, ou DBD, est le plus conservé chez les RN. Il permet la reconnaissance d'éléments de réponse spécifiques via deux motifs en doigt de zinc riches en cystéine (C-X2-C-X13-C-X2-C et C-X5-C-X9-C-X2-C) (Robinson-Rechavi *et al.*, 2003). Ces éléments de réponse sont composés de séquences hexamériques d'ADN répétées ou inversées et séparées par des séquences variant entre un à cinq nucléotides (Khorasanizadeh et Rastinejad, 2001). Le DBD offre aussi une surface de dimérisation via une structure appelée *D-box* (Pawlak *et al.*, 2012).

La région D, aussi appelée région charnière, est très peu conservée et a pour fonction principale de faciliter la rotation libre entre le DBD et le domaine de liaison au ligand (LBD - *Ligand binding domain*). Un signal de localisation nucléaire (NLS - *Nuclear localization signal*) contenu dans cette région participe à la régulation de la distribution subcellulaire des RN (Germain *et al.*, 2006).

Le domaine E, ou domaine de liaison au ligand, est constitué d'une pochette hydrophobe permettant la liaison d'un ligand. Une multitude de petites molécules lipophiles peuvent activer spécifiquement le LBD des RN, telles que les hormones stéroïdiennes, les phospholipides, les acides gras et les xénobiotiques (Pawlak *et al.*, 2012). Certains RN pour lesquels aucun ligand n'a encore été identifié sont considérés comme des récepteurs orphelins. Une deuxième région activatrice AF-2 est située dans le LBD. Contrairement à la région AF-1, l'activité de la région AF-2 est dépendante de la liaison du ligand au LBD. Celle-ci induit un changement de conformation dans le LBD, générant une pochette pouvant interagir avec le motif LXXLL présent sur une panoplie de co-activateurs transcriptionnels (Heery *et al.*, 1997). Le LBD, tout comme le DBD, contient une interface importante pour la dimérisation des récepteurs.

À l'extrémité C-terminale des RN se retrouve le domaine F. En raison de sa séquence hautement variable, la fonction exacte du domaine F reste encore à établir. Plusieurs RN n'ont notamment aucun domaine F. Néanmoins, la délétion de ce domaine chez les récepteurs qui le possèdent démontre dans certains cas son importance en lien avec diverses fonctions comme la dimérisation, l'activation et l'interaction avec différents corégulateurs (Patel et Skafar, 2015).

2.2 Activation et fonctions des récepteurs nucléaires

Avant l'établissement d'une nomenclature unifiée des RN en fonction de leur rapprochement phylogénétique, ceux-ci étaient regroupés en quatre classes selon leurs propriétés de liaison à l'ADN et de dimérisation (Mangelsdorf *et al.*, 1995). Bien que cette classification ait maintenant été remplacée, elle permet toujours d'illustrer les différences dans les modes d'activation et les fonctions de cette famille de facteurs de transcription. La plus grande distinction à ce niveau sépare les RN possédant un ligand et ceux pour lesquels aucun n'est encore connu (orphelins). Bien que les récepteurs orphelins possèdent

également un domaine de liaison au ligand (LBD), on n'y retrouve aucun ligand nécessaire à leur activation. Certains RN sont également considérés comme orphelins car la nature exacte de leur ligand n'est pas encore connue. Le LBD de ces récepteurs permet tout de même leur dimérisation et leur interaction avec divers corégulateurs transcriptionnels. Plusieurs RN orphelins peuvent s'associer à l'ADN et ainsi devenir actifs sous forme d'hétérodimères ou encore de monomères (Giguère, 1999 ; Bain *et al.*, 2007).

À l'exception des récepteurs orphelins, les RN sont typiquement connus pour être activés suite à leur liaison à un ligand spécifique. Ces ligands sont de petites molécules lipophiles capables de traverser les membranes plasmiques afin de lier leur récepteur intracellulaire. En absence de ligand, la région activatrice AF-2, située dans le domaine E des RN, prend une conformation ouverte. Sous cette conformation, cette région est accessible pour la liaison à différents corépresseurs transcriptionnels, menant à une répression de l'expression génique par les RN non liés à leur ligand (Sever et Glass, 2013). Les corépresseurs des récepteurs nucléaires NCoR1 et NCoR2 permettent la formation d'un complexe de répression en recrutant notamment les histones désacétylases HDAC1, HDAC3, HDAC4 et SIRT1 (Watson *et al.*, 2012). Un changement de conformation de la région AF-2 est induit suite à la liaison du ligand, délogeant les corépresseurs des RN. D'autre part, cette nouvelle conformation permet aussi la reconnaissance du motif LXXLL retrouvé dans la séquence de multiples coactivateurs, ce qui participe à la régulation positive de la transcription par les RN. Certains RN, tels le récepteur des androgènes (AR - NR3C4) et le récepteur des glucocorticoïdes (GR - NR3C1), sont localisés dans le cytosol dans un état inactif et ont besoin de la liaison au ligand pour transloquer au noyau. D'autres ont une localisation nucléaire constitutive, indépendante de la liaison au ligand (Germain *et al.*, 2006). Différents modes de dimérisation sont favorisés selon le type de RN. Certains d'entre eux sont des homodimères exclusifs, d'autres forment des hétérodimères principalement avec le récepteur nucléaire RXR (*Retinoid X receptor*), ou encore lient l'ADN directement sous forme de monomère. Les éléments de réponse des RN, aussi appelés HRE (*Hormone response elements*), peuvent être des répétitions directes (DR - *Direct repeats*), indirectes ou inversées de deux demi-sites d'ADN séparés par une séquence nucléotidique de taille variable (Giguère, 1999 ; Germain *et al.*, 2006). Il a été montré chez certains RN, tels RAR (*Retinoic acid receptor*) et RXR, que l'orientation et l'espacement des HRE reconnus contribuent à déterminer l'état d'activation ou de répression transcriptionnelle de leurs

gènes cibles (Rosenfeld *et al.*, 2006). Cette grande diversité de modes d'action des RN, possible malgré le fait qu'ils possèdent plusieurs éléments structuraux conservés, est associée à la multitude de fonctions impliquant les RN. Les récepteurs nucléaires sont notamment des régulateurs essentiels pour des processus biologiques tels le développement, le métabolisme, la différenciation et la prolifération cellulaire, la reproduction et l'apoptose (Gronemeyer *et al.*, 2004 ; Sever et Glass, 2013).

2.3 Isoformes des récepteurs nucléaires

Un autre niveau de complexité s'ajoute aux fonctions diverses des récepteurs nucléaires lorsqu'on considère que la plupart d'entre eux peuvent être exprimés en isoformes distinctes. Le récepteur de la progestérone (PR) (NR3C3) fut un des premiers RN pour lequel l'existence d'isoformes a été montrée (Kastner *et al.*, 1990). Les isoformes PR-A et PR-B possèdent des fonctions opposées, agissant respectivement en tant que répresseur et activateur transcriptionnels pour les mêmes gènes. Un des mécanismes proposés implique l'interaction préférentielle de l'isoforme PR-A avec le corépresseur NCoR2 (Giangrande et McDonnell, 1999).

Les isoformes des RN sont générées par différents mécanismes, incluant l'utilisation de promoteurs alternatifs, l'épissage alternatif et l'utilisation sélective de codons initiateurs. De façon globale, il est possible de les regrouper en quatre catégories en fonction des domaines structuraux qui sont modifiés (Laudet, 2006). Les isoformes les plus communément retrouvées présentent des variations au niveau du domaine N-terminal des RN. C'est le cas du récepteur PR, décrit ci-haut, ainsi que pour le récepteur RAR β , qui peut exprimer quatre isoformes possédant des extrémités N-terminales différentes. Un deuxième groupe d'isoformes touche le domaine de liaison à l'ADN de certains RN. Le récepteur PR peut également exprimer une troisième isoforme, PR-C, possédant un seul motif en doigt de zinc dans son DBD. Cette isoforme, ainsi incapable de lier l'ADN, peut cependant lier la progestérone et moduler l'activité de PR-A et PR-B par un mécanisme encore inconnu (Wei *et al.*, 1996). La troisième classe d'isoformes inclut les variants « dominants négatifs », pouvant lier l'ADN sans pouvoir transactiver leurs gènes cibles. Par exemple, le récepteur alpha des hormones thyroïdiennes (TR α) (NR1A1) est épissé alternativement en deux isoformes, TR α 1 et TR α 2, pouvant lier leurs éléments de réponse de la même manière. Le

variant TR α 2 ne peut cependant pas lier son ligand, et fonctionne ainsi comme un répresseur de la transcription en compétitionnant pour les sites de liaison avec l'isoforme active TR α 1 (Yang *et al.*, 1996). La dernière classe est composée d'isoformes possédant des insertions ou des délétions de courtes séquences dans leurs différents domaines. Le RN HNF4 α (*Hepatocyte nuclear factor 4 alpha*) exprime notamment une isoforme contenant une insertion de 10 acides aminés dans son domaine F (Chartier *et al.*, 1994). De manière générale, l'impact fonctionnel de ces courtes modifications est encore très peu compris. Cette diversité d'isoformes de RN pouvant être exprimées représente un mécanisme important permettant un niveau supplémentaire de régulation transcriptionnelle par ces récepteurs.

3. Le récepteur nucléaire HNF4 α

HNF4 α (*Hepatocyte Nuclear Factor 4 alpha*) (NR2A1) est un facteur de transcription ayant été identifié initialement en tant que régulateur de l'expression de gènes spécifiques au foie (Costa *et al.*, 1989). La protéine est clonée l'année suivante par le même groupe, et devient caractérisée en tant que récepteur nucléaire par homologie de séquence (Sladek *et al.*, 1990). Depuis, HNF4 α a été retrouvé également au niveau des reins, du pancréas, de l'estomac, de l'intestin grêle et du côlon (Tanaka *et al.*, 2006). Malgré tout, HNF4 α est principalement connu pour ses rôles dans le développement et le maintien de la fonction hépatocytaire, tels la différenciation, le transport et le métabolisme des lipides, ainsi que la détoxification des xénobiotiques. Il a d'ailleurs été observé que HNF4 α se lie à près de la moitié des gènes transcrits activement dans le foie, démontrant son rôle central dans le développement et la fonction hépatocytaire (Odom *et al.*, 2004).

3.1 Structure et domaines de HNF4 α

HNF4 α est une protéine appartenant à la superfamille des récepteurs nucléaires, dû au fait qu'elle possède les domaines fonctionnels caractéristiques de cette classe de facteurs de transcription, identifiés classiquement de A à F et décrits en détail dans la section 2.1 (Figure 1). Il n'existe présentement aucune structure tridimensionnelle complète de HNF4 α , en raison du faible niveau d'organisation structurale des extrémités N- et C-terminales. La

forme la plus complète ayant pu être cristallisée à ce jour regroupe les domaines C, D et E liés à l'ADN sous une forme d'homodimère (Figure 2) (Chandra *et al.*, 2013).

Le domaine A/B de HNF4 α , situé à l'extrémité N-terminale, contient une région d'activation AF-1 (*Activation function*) permettant son interaction avec divers corégulateurs et membres de la machinerie de transcription générale (Tableau 1). Le domaine C, aussi appelé domaine de liaison à l'ADN, permet la reconnaissance d'éléments de réponse spécifiques situés dans les régions régulatrices de ses gènes cibles. HNF4 α reconnaît des sites DR1 (*Direct Repeat 1*), composés de deux demi-sites hexamériques répétés et séparés par un nucléotide, typiquement une adénosine. HNF4 α reconnaît également les répétitions directes séparées par deux nucléotides (DR2), mais avec une plus faible spécificité (Jiang et Sladek, 1997). La séquence consensus de ses demi-sites, AGGTCA, est partagée par la plupart des RN non stéroïdiens (Fang *et al.*, 2012). La liaison de ces sites par HNF4 α est possible dû à la présence de deux motifs en doigts de zinc au sein du DBD, et celui-ci offre aussi une interface de dimérisation du récepteur (Jiang et Sladek, 1997).

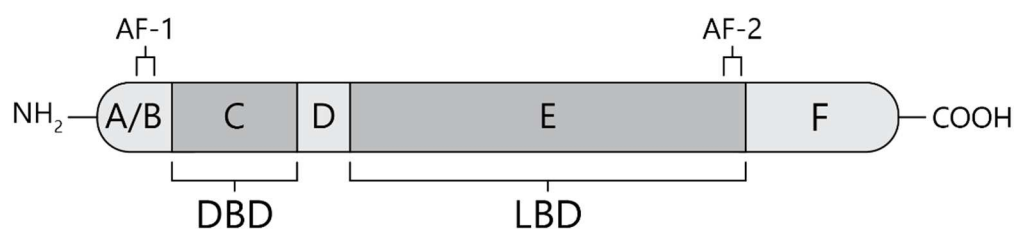


Figure 1. Les domaines fonctionnels du récepteur nucléaire HNF4 α . Représentation linéaire de la structure de HNF4 α et de ses différents domaines. Les domaines de liaison à l'ADN (DBD) et du ligand (LBD) (domaines C et E, respectivement) sont les plus conservés par rapport aux autres récepteurs nucléaires. Le domaine A/B, situé en N-terminal, contient une région activatrice AF-1. Une seconde région activatrice, AF-2, est située à la toute fin du domaine E. Le domaine F, situé à l'extrémité C-terminale, est considéré comme un domaine répresseur.

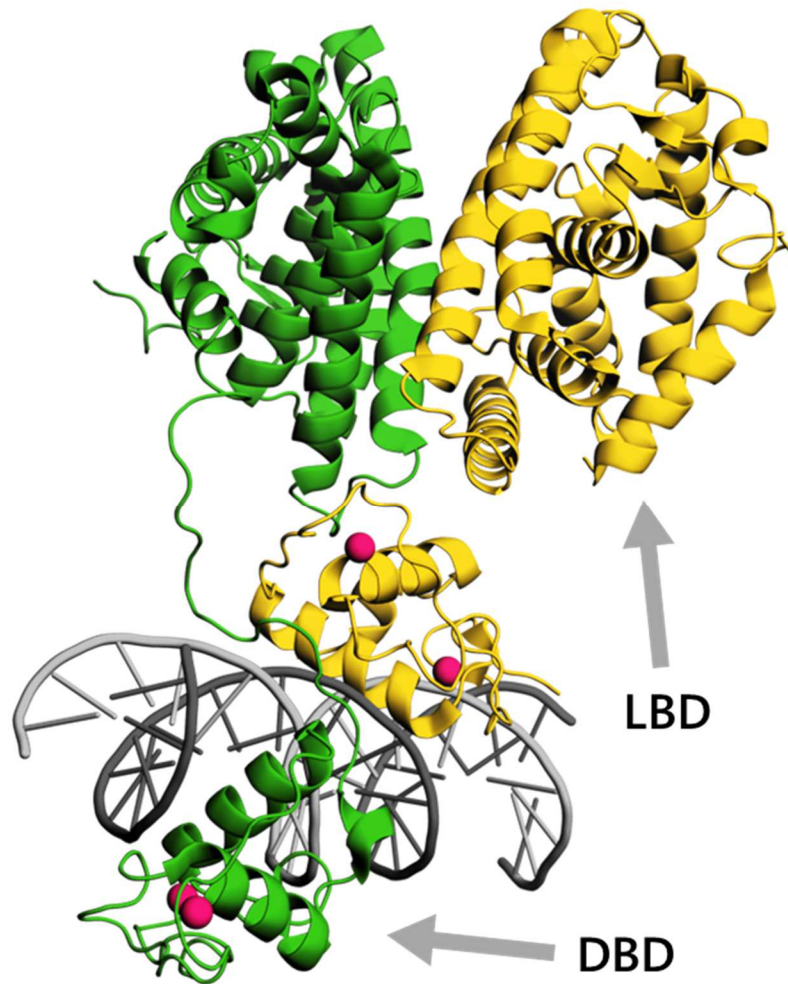


Figure 2. HNF4a lie l'ADN sous forme d'homodimère. Cristallographie d'une forme partielle du récepteur nucléaire HNF4a lié à son élément de réponse DR1 sous forme d'homodimère (vert et jaune). Le domaine de liaison à l'ADN (DBD) coordonne la liaison de deux atomes de Zn^{2+} (en rose) via leurs motifs en doigts de zinc et offre une surface d'homodimérisation. Le domaine de liaison au ligand participe aussi à l'homodimérisation du récepteur. Adapté de Chandra *et al.*, 2013. PDB ID : 4IQR.

Le domaine E, aussi appelé domaine de liaison au ligand, est une autre région essentielle à l'homodimérisation de HNF4α. Il contient de plus la région activatrice AF-2 dans son extrémité C-terminale, importante pour la liaison de plusieurs corégulateurs transcriptionnels (Tableau 1). Bien que la présence d'une pochette de liaison pour un ligand ait été identifiée par cristallographie, la nature et la fonction d'un ligand pour HNF4α demeurent toujours un sujet de controverse, et sont discutées de façon plus approfondie dans la section suivante (Dhe-Paganon *et al.*, 2002). À l'extrémité C-terminale du récepteur se retrouve le domaine F, qui est absent chez plusieurs RN. Le domaine F de HNF4α est particulièrement long et est considéré comme un domaine répresseur, puisque sa délétion augmente significativement l'activité du récepteur (Sladek *et al.*, 1999). Cette fonction répressive a été associée à une courte région de 14 acides aminés au sein de ce domaine. Cela représente un mécanisme unique chez les récepteurs nucléaires, qui demeure toujours peu compris dû au fait que la structure complète de HNF4α n'ait jamais été élucidée (Hadzopoulou-Cladaras *et al.*, 1997 ; Iyemere *et al.*, 1998).

Tableau 1 : Les régions activatrices AF-1 et AF-2 coordonnent les interactions de HNF4α avec divers corégulateurs transcriptionnels. Liste non exhaustive des corégulateurs interagissant de manière spécifique ou non avec les régions AF-1 et AF-2 de HNF4α, distingués en coactivateurs, corépresseurs et membres du complexe de transcription basale.

	AF-1	AF-2
Coactivateurs transcriptionnels	CBP ¹ , p300 ² , NCOA2 ³ , MED14 ⁴ , PC4 ⁵ , SMAD3/4 ⁶ , TADA2L ⁵	CBP, p300, NCOA1 ³ /2/3 ⁷ , MED1 ⁸ , PGC-1α ⁹
Corépresseurs transcriptionnels		NCOR2 ² , SHP ⁷ , SMILE ¹⁰
Complexe de transcription basale	TBP ⁵ , TFIIB ⁵ , TAF6 ⁵ , TAF9 ⁵ , GTF2H1 ⁵	

¹ Dell et Hadzopoulou-Cladaras, 1999.

² Torres-Padilla *et al.*, 2002.

³ Wang *et al.*, 1998.

⁴ Malik *et al.*, 2002.

⁵ Green *et al.*, 1998.

⁶ Chou *et al.*, 2003.

⁷ Lee *et al.*, 2000.

⁸ Maeda *et al.*, 2002.

⁹ Yoon *et al.*, 2001.

¹⁰ Xie *et al.*, 2009.

3.2 Mode d'action de HNF4α

HNF4α a pendant longtemps été considéré comme un récepteur nucléaire orphelin, puisqu'aucun ligand n'était connu pour moduler son activité. Contrairement à d'autres types de récepteurs nucléaires, HNF4α se retrouve localisé de façon constitutive au noyau et ne nécessite pas l'ajout exogène d'un ligand pour pouvoir s'homodimériser et interagir avec les éléments de réponse de ses gènes cibles. La cristallographie du LBD a permis initialement d'observer la présence d'acides gras à composition diverse liés au niveau de la pochette de liaison du ligand de HNF4α (Dhe-Paganon *et al.*, 2002). Des études subséquentes ont identifié l'acide linoléique, un acide gras polyinsaturé à longue chaîne (C18:2ω6), comme la molécule se liant préférentiellement à son LBD (Yuan *et al.*, 2009). Cette liaison est réversible, et ne module pas l'activité transcriptionnelle de HNF4α. La nature du ligand de HNF4α est donc toujours controversée, puisque l'acide linoléique est présent de façon endogène et ne semble pas nécessaire à l'activité de son récepteur, contrairement au mode d'action typique des récepteurs nucléaires nécessitant la liaison à leur ligand. HNF4α est considéré comme un homodimère exclusif, retrouvé de façon stable sous cette forme en solution, et il doit être sous cette forme afin de pouvoir lier l'ADN. Les récepteurs nucléaires RXRα/β/γ et RARα, connus pour leur capacité à former des hétérodimères avec plusieurs RN, ne s'assemblent pas en hétérodimères avec HNF4α (Jiang *et al.*, 1995 ; Lee et Privalsky, 2005). Une fois lié à ses différents éléments de réponse DR1 et DR2, HNF4α régule l'expression de ses gènes cibles par deux mécanismes principaux (Figure 3).

D'une part, son interaction avec différents coactivateurs facilite le recrutement du PIC au promoteur de ses gènes cibles, et mène ainsi à l'expression de ceux-ci (Figure 3A). Notamment, HNF4α interagit directement avec différents membres de la machinerie de transcription générale, incluant les GTF TBP et TFIIB, ainsi que certaines sous-unités de GTF comme TAF6, TAF9 et GTF2H1 (Green *et al.*, 1998). Cette même étude a montré son interaction avec la protéine PC4, qui coordonne l'association entre différents activateurs et le complexe de transcription basale (Green *et al.*, 1998 ; Malik *et al.*, 1998). HNF4α interagit aussi avec certaines sous-unités du complexe coactivateur *Mediator*, telles que MED1 et MED14, via ses régions d'activation fonctionnelle AF-1 et AF-2, respectivement (Maeda *et al.*, 2002 ; Malik *et al.*, 2002). MED25 a également été démontrée comme étant une sous-unité nécessaire au recrutement du complexe *Mediator* par HNF4α (Rana *et al.*, 2011).

De façon plus indirecte, HNF4 α favorise le recrutement du PIC par l'entremise de son interaction avec d'autres coactivateurs capables d'induire une configuration plus relâchée de la chromatine. L'histone acétyltransférase p300 est un de ces coactivateurs, permettant l'augmentation de la transcription du FT HNF1 α par HNF4 α (Eeckhoutte *et al.*, 2004). Une autre histone acétyltransférase, CBP (*CREB-binding protein*), interagit à la fois avec la région AF-1 et la région AF-2 de HNF4 α via ses extrémités N- et C-terminales, respectivement (Dell et Hadzopoulou-Cladaras, 1999). HNF4 α interagit également avec l'adaptateur transcriptionnel TADA2L, qui fait partie du complexe histone acétyltransférase ATAC (*Ada2a-containing HAT complex*) (Green *et al.*, 1998). Les membres de la famille p160 des coactivateurs des récepteurs nucléaires NCoA-1 et NCoA-3 possèdent également une activité HAT, qui est cependant beaucoup plus faible que celle provenant des enzymes décrites précédemment. La fonction principale de ces coactivateurs pour HNF4 α , ainsi que du dernier membre de la famille, NCoA-2, est plutôt d'agir en tant qu'intermédiaires pour le recrutement d'autres coactivateurs. Les coactivateurs p160 interagissent donc avec HNF4 α sous forme de complexes avec d'autres protéines telles que les histones acétyltransférases CBP, p300 et KAT2B (*Lysine acetyltransferase 2B*), ainsi que les histones méthyltransférases CARM1 (*Coactivator-associated arginine methyltransferase 1*) et PRMT1 (*Protein arginine N-methyltransferase 1*) (Xu et Li, 2003). PRMT1 a également le potentiel d'interagir directement avec HNF4 α (Barrero et Malik, 2006). CBP et p300 peuvent à leur tour recruter le complexe de remodelage de la chromatine dépendant de l'ATP SWI/SNF au niveau de la région liée par HNF4 α (Xu et Li, 2003).

De manière opposée, HNF4 α est capable de bloquer l'expression de certains de ses gènes cibles en recrutant différents corépresseurs (Figure 3B). NCoR2, connu pour être un corépresseur de plusieurs récepteurs nucléaires, peut compétitionner avec les coactivateurs CBP, p300 et NCoA-2 pour les sites de liaison retrouvés dans la structure de HNF4 α de façon mutuellement exclusive (Ruse *et al.*, 2002). L'association de NCoR2 en complexe avec différentes protéines incluant HDAC3 et HDAC4 mène également à la désacétylation des histones et ainsi à une compaction de la chromatine suite à son recrutement par HNF4 α (Torres-Padilla *et al.*, 2002). Le récepteur nucléaire SHP (NR0B2) est particulier puisqu'il ne possède pas de DBD, et agit plutôt comme un corépresseur pour plusieurs autres RN. Il interagit avec la région d'activation transcriptionnelle AF-2 de HNF4 α , causant une compétition pour la liaison de cette région avec le coactivateur NCoA-2 (Lee *et al.*, 2000).

De la même manière, le corépresseur SMILE peut antagoniser la liaison du coactivateur PGC-1 α en interagissant avec le LBD de HNF4 α , et recrute HDAC1, HDAC3 et HDAC4 comme un second mécanisme de répression du récepteur (Xie *et al.*, 2009).

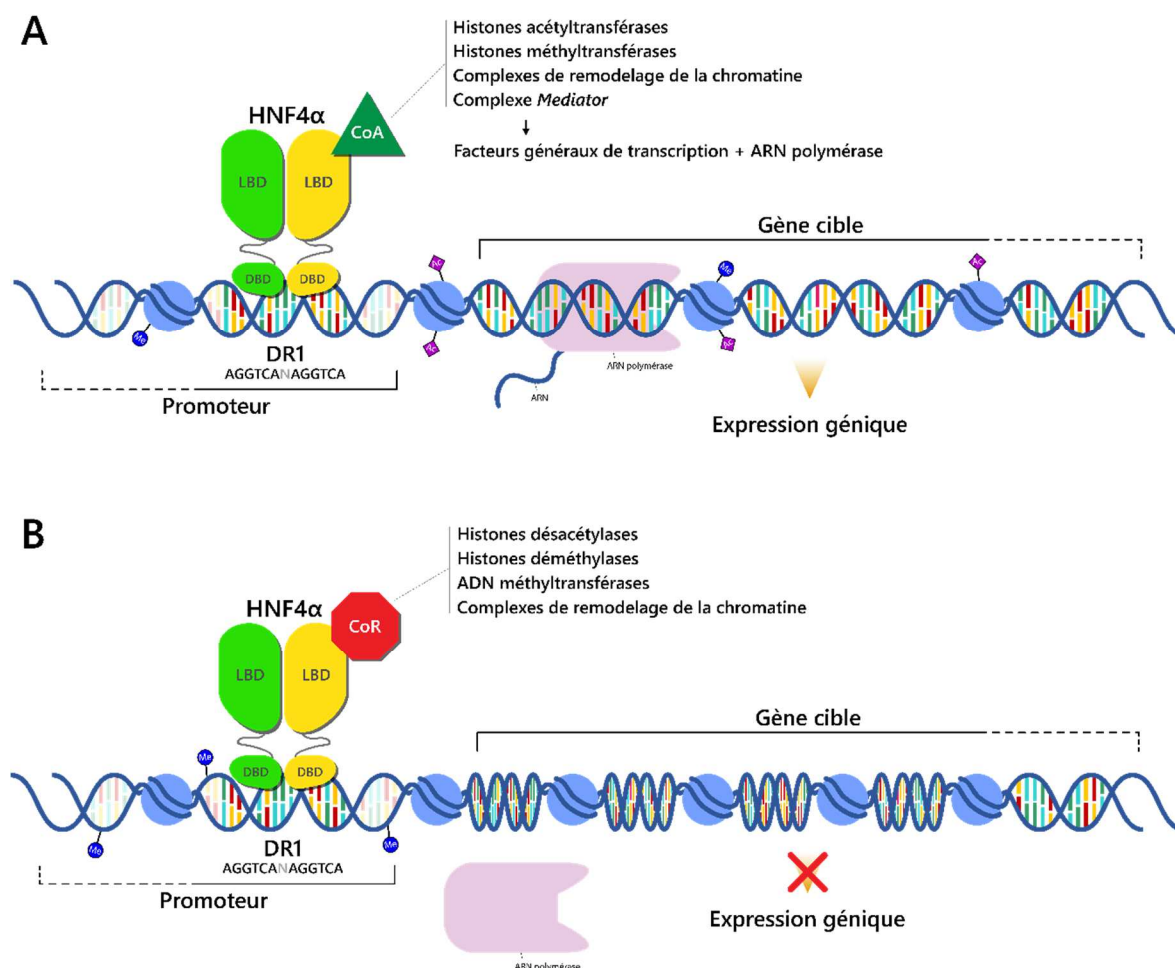


Figure 3. Les coactivateurs et corépresseurs interagissant avec HNF4 α régulent son activité transcriptionnelle. Le récepteur nucléaire HNF4 α reconnaît son élément de réponse consensus DR1 et peut interagir avec différents corégulateurs afin d'activer ou de réprimer l'expression de ses gènes cibles. **(A)** Activation transcriptionnelle par HNF4 α via son interaction avec différents coactivateurs. Ceux-ci ont pour fonction de recruter le complexe de l'ARN polymérase II ou de favoriser un état relâché de la chromatine. **(B)** Répression transcriptionnelle par HNF4 α suite à son interaction avec certains corépresseurs, dont la fonction est de compétitionner la liaison des coactivateurs et de provoquer un état plus compacté de la chromatine, empêchant le recrutement de la PolII.

Finalement, une panoplie d'autres facteurs de transcription interagissent avec HNF4α dans certains contextes pour moduler son activité transcriptionnelle par différents mécanismes. HNF1α (*Hepatocyte nuclear factor 1 alpha*) est un FT capable de lier directement la région AF-2 de HNF4α, réprimant ainsi son activité transcriptionnelle (Ktistaki et Talianidis, 1997). Selon un mécanisme similaire, les isoformes 1a et 1c du FT SREBP (*Sterol regulatory element-binding protein*) compétitionnent avec le coactivateur PGC-1α pour la liaison du LBD de HNF4α (Yamamoto *et al.*, 2004). La voie de signalisation NF-κB (*Nuclear factor kappa-light-chain-enhancer of activated B cells*), pouvant être activée notamment par la cytokine TNFα (*Tumour necrosis factor alpha*), régule négativement la transcription du gène *APOC3* par HNF4α en inhibant sa liaison à son élément de réponse ainsi que l'activité des régions AF-1 et AF-2 (Nikolaidou-Neokosmidou *et al.*, 2006). Les récepteurs nucléaires COUP-TFI/II (*Chicken ovalbumin upstream promoter transcription factor I/II*) (NR2F1 et NR2F2) reconnaissent plusieurs éléments de réponse également reconnus par HNF4α, incluant le site C3P du gène *APOC3*, et régulent donc négativement son activité transcriptionnelle en réduisant sa liaison à l'ADN (Mietus-Snyder *et al.*, 1992). Dans un autre contexte, comme au promoteur du gène *HNF1A*, ces mêmes RN ne compétitionnent pas pour la liaison de HNF4α à son élément de réponse, mais augmentent plutôt son activité transcriptionnelle en interagissant avec son LBD (Ktistaki et Talianidis, 1997).

Dans l'ensemble, ces exemples de coactivateurs, corépresseurs et autres facteurs de transcription interagissant avec HNF4α illustrent la complexité entourant la régulation de l'activité transcriptionnelle de celui-ci. Plusieurs mécanismes sont utilisés par ces acteurs transcriptionnels pour moduler sa fonction. Afin de mieux comprendre les rôles que joue HNF4α dans divers processus biologiques, l'intégration de ces mécanismes de régulation doit évidemment tenir compte du contexte cellulaire dans lequel ces études d'interaction ont été réalisées.

3.3 Processus biologiques régulés par HNF4α

HNF4α est impliqué dans une foule de fonctions biologiques associées aux tissus qui l'expriment tels le foie, le pancréas et l'intestin. Différentes expériences portant sur l'invalidation de HNF4α ont d'abord permis de démontrer son implication dans plusieurs processus développementaux, comme l'embryogénèse et la morphogénèse. L'invalidation

totale du gène *HNF4A* chez la souris est létale au stade embryonnaire, due à des défauts de gastrulation et à une mort cellulaire au niveau de l'endoderme viscéral (Chen *et al.*, 1994). La complémentation de ce modèle avec un endoderme viscéral normal permet de prévenir la létalité embryonnaire et d'étudier l'impact de la perte de HNF4α au niveau du foie. Bien que le développement précoce du foie ne soit pas affecté par l'absence de HNF4α, d'importantes perturbations dans la différenciation et la fonction des hépatocytes se manifestent dans ces conditions (Li *et al.*, 2000). L'invalidation de HNF4α de manière spécifique à l'épithélium du côlon cause des défauts au niveau de la formation des cryptes, de la maturation des cellules à mucus et de l'expression de plusieurs gènes importants pour la fonction du côlon (Garrison *et al.*, 2006). Les souris invalidées pour HNF4α spécifiquement au niveau des cellules bêta pancréatiques ne présentent pas de défauts dans l'architecture des îlots pancréatiques, mais démontrent une hyperinsulinémie et une intolérance au glucose (Gupta *et al.*, 2005).

Chez l'adulte, HNF4α est impliqué dans la régulation de l'homéostasie et du métabolisme de plusieurs composantes biologiques. L'invalidation conditionnelle de HNF4α spécifique au foie a permis de montrer l'implication de ce FT dans l'expression de plusieurs gènes essentiels pour le métabolisme et le transport des lipides et du cholestérol (Hayhurst *et al.*, 2001). HNF4α contrôle de plus le métabolisme des xénobiotiques et des acides biliaires en activant l'expression de plusieurs membres de la famille des cytochromes P450 (Jover *et al.*, 2001 ; Inoue *et al.*, 2006). L'homéostasie du glucose est également régulée par HNF4α, notamment via l'expression de la protéine de transport du glucose SLC2A2 (*Solute carrier family 2 member 2*), différentes enzymes glycolytiques, ainsi que certaines protéines régulant la sécrétion insulinaire (Stoffel et Duncan, 1997 ; Wang *et al.*, 2000). Outre ces fonctions, HNF4α a été associé à la fois à la différenciation et la prolifération cellulaire, dépendamment du contexte étudié. Diverses évidences montrent son implication en tant que régulateur essentiel de la maintenance d'un état différencié des hépatocytes (Hayhurst *et al.*, 2001 ; DeLaForest *et al.*, 2011). Il participe aussi à la différenciation des cellules épithéliales intestinales via la régulation positive des facteurs de transcription HNF1α et CDX-2 (Lussier *et al.*, 2008 ; Boyd *et al.*, 2009). De façon reliée à son rôle dans la différenciation, HNF4α est aussi associé à un ralentissement de la prolifération dans les lignées cellulaires HEK 293, INS-1 et HCT 116 (Lucas *et al.*, 2005 ; Erdmann *et al.*, 2007 ; Hwang-Verslues et Sladek, 2008). L'effet observé dans les cellules HCT 116 est expliqué en

partie par la régulation positive de l'inhibiteur du cycle cellulaire p21 par HNF4 α , qui compétitionne pour le contrôle de ce gène avec le FT c-Myc. Au niveau de l'épithélium intestinal, la régulation négative de la prolifération par HNF4 α semble passer par l'inhibition de la voie Wnt/ β -caténine (Cattin *et al.*, 2009). Par opposition à ces observations, l'inhibition de HNF4 α dans les lignées de cancer colorectal HT-29 et Caco-2 cause une diminution de la croissance et de la prolifération cellulaire (Schwartz *et al.*, 2009). De plus, l'expression de HNF4 α au niveau des cellules β pancréatiques est requise pour engendrer l'augmentation de la prolifération en réponse à un stress métabolique (Gupta *et al.*, 2007).

3.4 Régulation de HNF4 α

Malgré le fait que HNF4 α soit activé constitutivement en absence de ligand, plusieurs mécanismes de régulation contrôlent son activité transcriptionnelle. Le gène *HNF4A* est régulé par un promoteur proximal P1 et un promoteur distal P2 (46 000 pb en amont) (Thomas *et al.*, 2001). L'utilisation alternative de ces deux promoteurs mène à l'expression possible de deux classes d'isoformes, aussi appelées P1 et P2. Les différences de fonctions et d'expression impliquant ces isoformes représentent un premier niveau de régulation de l'activité transcriptionnelle de HNF4 α , et sont présentées en détail dans la prochaine section de ce chapitre. L'expression de HNF4 α est finement régulée par l'action de plusieurs FT au niveau de ses deux promoteurs. Les FT de la famille HNF (*Hepatocyte nuclear factors*), divisés en quatre familles (HNF1, FOXA/HNF3, HNF4 et ONECUT/HNF6), forment d'importants réseaux de régulation croisée. Au stade embryonnaire, HNF1 β (*Hepatocyte nuclear factor 1 beta*) et GATA6 (*GATA binding protein 6*) coopèrent pour activer la transcription de *HNF4A* (Lau *et al.*, 2018). Le FT FOXA2 (*Forkhead box A2*) est également un activateur de l'expression de HNF4 α , alors que FOXA1 (*Forkhead box A1*) antagonise cette activation en compétitionnant pour les mêmes sites de liaison (Duncan *et al.*, 1998). Dans le foie adulte, HNF1 α remplace HNF1 β en tant qu'activateur transcriptionnel de HNF4 α , et ce de façon synergique avec GATA6. Dans le même contexte, il a été montré que COUP-TFII réprimait l'expression de HNF4 α en compétitionnant avec l'activateur hétérodimérique RXR α -RAR α pour le même site DR1 dans le promoteur du gène *HNF4A* (Hatzis et Talianidis, 2001). Au niveau du pancréas chez l'adulte, l'expression de HNF4 α est contrôlée directement par HNF1 α (Boj *et al.*, 2001).

En plus d'un contrôle important au niveau de son expression par plusieurs FT, plusieurs caractéristiques fonctionnelles de HNF4α sont modulées par une multitude de modifications post-traductionnelles (MPT). L'acétylation de la lysine en position 458 du domaine F de HNF4α réduit son activité transcriptionnelle (Yokoyama *et al.*, 2011). Un autre mécanisme de régulation post-traductionnelle implique la protéine kinase A (PKA). Celle-ci reconnaît un site consensus RRSS dans la région *A-box* du DBD de HNF4α, et inhibe fortement sa liaison à l'ADN par phosphorylation (Viollet *et al.*, 1997). La localisation nucléaire de HNF4α est modulée par plusieurs MPT, telles que son acétylation par CBP (Soutoglou *et al.*, 2000) et la phosphorylation de sa sérine en position 78 par la protéine kinase C (PKC) (Sun *et al.*, 2007). Finalement, la phosphorylation de sa sérine en position 304 par AMPK (*5'-AMP-activated protein kinase*) peut réduire à la fois sa capacité d'homodimérisation et sa stabilité (Hong *et al.*, 2003 ; Leclerc *et al.*, 2001).

4. Les isoformes de HNF4α

4.1 Expression différentielle des classes d'isoformes P1- et P2-HNF4α

De façon globale, les isoformes de HNF4α peuvent être regroupées en deux classes, P1 et P2, en fonction du promoteur utilisé pour leur expression. L'expression de ces classes d'isoformes est finement contrôlée selon le contexte spatial et au cours du développement. L'analyse immunohistochimique de plusieurs tissus humains révèle une distribution variable de l'expression des isoformes P1 et P2 (Tableau 2) (Tanaka *et al.*, 2006). Le foie et le rein sont démontrés pour exprimer en grande majorité les isoformes de la classe P1, alors que le pancréas et l'estomac sont considérés comme spécifiques pour les isoformes P2. Un modèle de souris exprimant uniquement les isoformes P1 présente un phénotype d'intolérance au glucose, démontrant l'implication spécifique de P2 dans la fonction du pancréas. Un autre modèle de souris exprimant seulement P2-HNF4α montre une dyslipidémie ainsi qu'une stéatose hépatique, soulignant l'importance de l'expression des isoformes P1 au niveau du foie (Briangon et Weiss, 2006). L'intestin, autant au niveau de l'intestin grêle que du côlon, est le seul organe adulte exprimant à la fois les isoformes P1 et P2. Les niveaux d'expression de ces classes d'isoformes ont été modulés dans le côlon, par interférence à l'ARN dans une lignée de cancer colorectal et par échange d'exons chez

la souris, afin d'analyser les différences fonctionnelles entre P1 et P2. Ces études ont révélé des disparités importantes, les isoformes P1 étant impliquées dans la régulation de la différenciation cellulaire et du métabolisme, alors que les isoformes P2 étaient associées à la prolifération cellulaire et la progression du cancer (Babeu *et al.*, 2018 ; Chellappa *et al.*, 2016).

La balance d'expression entre les isoformes P1 et P2 est modulée au cours du développement par différents mécanismes de régulation transcriptionnelle. Au niveau du foie murin, HNF4 α est exprimé majoritairement sous sa forme P2 au stade foetal, et la classe P1 devient progressivement la forme prédominante au cours du développement (Torres-Padilla *et al.*, 2001). Le mécanisme derrière ce changement demeure encore à préciser, mais il semble que le promoteur P2 puisse être activé directement par les FT HNF1 β , HNF6 (*Hepatocyte nuclear factor 6*) et OC2 (*One cut domain family member 2*) dans le foie embryonnaire (Kymizi *et al.*, 2006 ; Briançon *et al.*, 2004). L'augmentation progressive de l'expression des isoformes P1 durant le développement du foie supprime directement cette activation du promoteur P2, menant à un foie adulte exprimant exclusivement P1-HNF4 α (Briançon *et al.*, 2004). Il a également été montré que l'exposition au glucocorticoïde synthétique dexaméthasone pousse la balance d'expression vers le promoteur P1 durant le stade foetal chez le rat (Nyirenda *et al.*, 2006). Au niveau du pancréas humain, jusqu'à 23 % des transcrits de HNF4 α proviennent du promoteur P1 au stade foetal, alors que l'adulte exprime uniquement la classe P2 (Harries *et al.*, 2008). Les FT HNF1 α , HNF1 β et PDX1 (*Pancreas/duodenum homeobox protein 1*) régulent positivement le gène *HNF4A* via son promoteur P2 dans ce tissu (Thomas *et al.*, 2001 ; Ferrer, 2002). Il a toutefois été rapporté par un groupe que les isoformes P1 pouvaient aussi être retrouvées dans les îlots pancréatiques chez l'adulte à un niveau comparable aux isoformes P2 (Eeckhoutte *et al.*, 2003).

Tableau 2 : L'expression des classes d'isoformes P1- et P2-HNF4a est compartimentée selon les tissus. Tableau résumant la distribution des classes d'isoformes de HNF4a exprimées dans différents tissus humains chez l'adulte (Tanaka *et al.*, 2006).

Tissu (adulte)	P1	P2
Foie	+	-
Pancréas	-	+
Rein	+	-
Estomac	-	+
Intestin grêle	+	+
Côlon	+	+

4.2 Identification des isoformes de HNF4a

Au niveau structurel, l'unique différence entre les classes P1- et P2-HNF4a se situe au niveau du domaine A/B à l'extrémité N-terminale des isoformes (Figure 4). Cette variation dans leur séquence fait en sorte que la région activatrice AF-1, qui permet l'interaction de HNF4a avec différents corégulateurs, est présente seulement chez les isoformes de la classe P1.

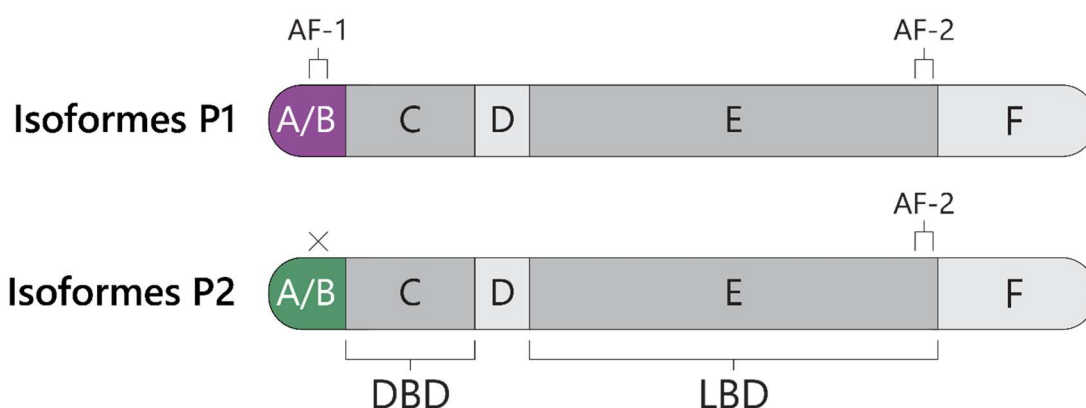


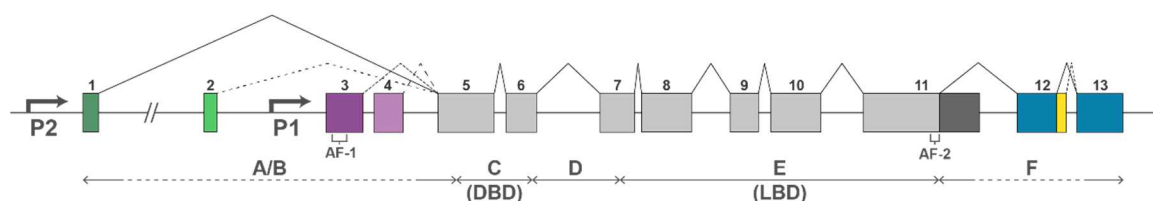
Figure 4 : Les classes d'isoformes P1- et P2-HNF4a varient uniquement au niveau de leur domaine A/B. Représentation générale de la structure linéaire et des domaines des isoformes P1 et P2. La majorité de la séquence des isoformes P1 et P2 est la même, avec une identité de séquence d'environ 93 %. Seul le domaine A/B varie entre les deux classes, faisant en sorte que la région activatrice AF-1 est retrouvée uniquement chez les isoformes P1.

Cette distinction ne s'applique cependant pas à l'ensemble des isoformes composant les classes P1 et P2. À travers la littérature entourant les isoformes de HNF4α, la terminologie d'isoformes P1 et P2 réfère uniquement aux premières isoformes ayant été identifiées et caractérisées plus en détail, et qui sont considérées comme canoniques dans cet ouvrage. Ces isoformes canoniques sont α1, α2 et α3 pour la classe P1 et α7, α8 et α9 pour la classe P2. Les autres isoformes, qui seront ici désignées comme non canoniques, ne partagent pas le même domaine A/B que les autres isoformes qui composent leur classe, et sont seulement regroupées dû au fait que leur expression découle d'un promoteur commun.

Depuis l'identification initiale du récepteur nucléaire HNF4α dans le foie, qui s'est avérée plus tard être l'isoforme α1, un total de 12 isoformes ont été montrées pour être exprimées ou dont l'existence a été prédite en tenant compte du mécanisme d'épissage alternatif (Figure 5). La numérotation des différents exons du gène *HNF4A* a été corrigée pour des besoins de simplification dans cet ouvrage, et ne correspond donc pas aux différentes identifications utilisées dans la littérature. Outre les trois isoformes dites canoniques composant chaque classe, on retrouve les isoformes P1 α4, α5 et α6 ainsi que les isoformes P2 α10, α11 et α12. Les isoformes α1 et α2 furent les premières identifiées chez l'humain au niveau du foie. L'utilisation d'un site d'épissage donneur alternatif cause l'insertion d'une séquence additionnelle de 10 acides aminés dans le domaine F de HNF4α2 (Chartier *et al.*, 1994). L'isoforme α3, identifiée un peu plus tard, partage le domaine A/B de α1 et α2, mais la séquence de son domaine F est complètement différente en raison de l'utilisation d'un autre site d'épissage donneur alternatif (Kritis *et al.*, 1996). Les isoformes α7, α8 et α9 partagent respectivement la même structure au niveau de leur extrémité C-terminale que les isoformes α1, α2 et α3, et varient seulement au niveau de leur domaine A/B suite à l'utilisation du promoteur P2 (Thomas *et al.*, 2001 ; Torres-Padilla *et al.*, 2001 ; Hansen *et al.*, 2002). Le clonage initial de l'isoforme α4 montrait qu'elle contenait dans son domaine A/B une séquence additionnelle de 30 acides aminés par rapport à l'isoforme α2 (Drewes *et al.*, 1996). Cette observation est toutefois fausse, puisque la séquence de α4 publiée à l'origine contenait une erreur. L'ajout de l'exon 4 à l'exon 3 par épissage alternatif mène en réalité à l'apparition d'un codon stop prématuré. L'utilisation d'un codon initiateur alternatif dans ce contexte permet toutefois l'expression de l'isoforme α4 avec un domaine A/B complètement différent de celui des isoformes α1, α2 et α3, contrairement à ce qui

avait été publié initialement. Une explication plus détaillée entourant la séquence réelle de HNF4 α 4 est présentée dans l'annexe B. Suite au clonage de α 4, l'existence des isoformes α 5 et α 6 fut prédite en considérant l'épissage alternatif du domaine F connu pour les isoformes α 1/2/3 (Furuta *et al.*, 1997). Les isoformes α 10, α 11 et α 12 sont les dernières à avoir été identifiées, et sont régulées par le promoteur distal P2 (Huang *et al.*, 2009). De façon étonnante, l'ajout de l'exon 2 à l'exon 1 par épissage alternatif provoque la même situation que pour les isoformes α 4/5/6, en menant à l'apparition d'un codon stop prématuré. L'utilisation d'un codon initiateur alternatif fait en sorte que les isoformes α 10/11/12 possèdent un domaine A/B entièrement distinct de celui des isoformes α 7/8/9, mêmes si elles sont toutes régulées par le promoteur P2. Le tableau 10 (Annexe C) récapitule l'identification des isoformes de HNF4 α à travers les bases de données RefSeq, Uniprot et Ensembl.

Gène *HNF4A*



Isoformes de HNF4 α

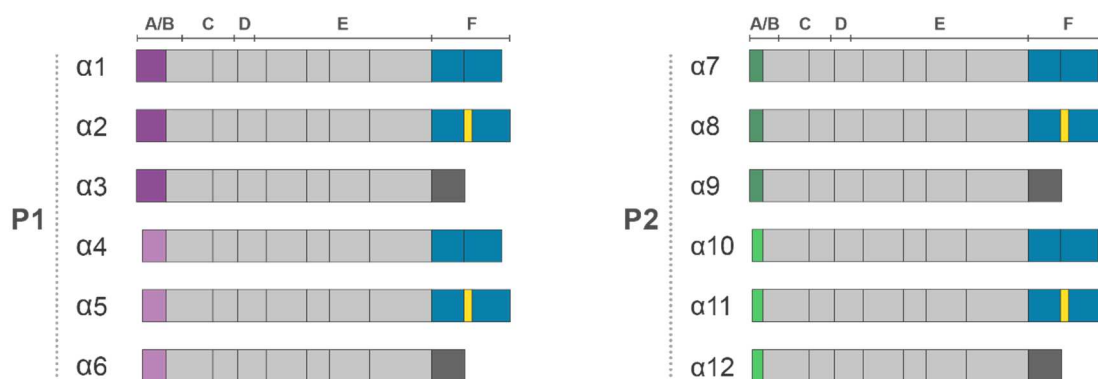


Figure 5 : Le gène *HNF4A* peut mener à l'expression de 12 isoformes, par l'utilisation de promoteurs alternatifs et par épissage alternatif. Le promoteur P1 proximal peut produire six isoformes (α 1- α 6), et autant d'isoformes provenant du promoteur P2 distal sont possibles (α 7- α 12). Les deux promoteurs du gène *HNF4A* sont séparés par environ 46 000 pb.

4.3 Activité transcriptionnelle des isoformes de HNF4α

La plupart des études qui se sont intéressées aux isoformes de HNF4α considéraient seulement les classes P1 et P2 de manière globale. Comme mentionné précédemment, les différences fonctionnelles démontrées entre ces classes ne s'appliquent qu'aux isoformes canoniques (α1, α2 et α3 pour P1 ; α7, α8 et α9 pour P2), alors que les autres isoformes composant ces classes ont été très peu étudiées au niveau de leur rôle dans la régulation transcriptionnelle. De manière plus spécifique, certaines études visaient à comparer la capacité de transactivation entre certaines isoformes, ainsi que les interactions variables au niveau de leurs corégulateurs.

La différence la mieux caractérisée implique le domaine A/B, qui contient une région activatrice AF-1 uniquement chez les isoformes P1 canoniques. L'introduction des isoformes α2 et α8 dans les cellules HCT 116 a permis d'étudier l'impact fonctionnel de cette différence structurelle. Dans ce contexte, HNF4α2 régule positivement plusieurs gènes impliqués dans l'arrêt de croissance et l'apoptose, alors que HNF4α8 est associée à des fonctions prolifératives et anti-apoptotiques (Vuong *et al.*, 2015). L'interaction préférentielle de ces isoformes avec certains régulateurs transcriptionnels pourrait expliquer ces fonctions distinctes. En effet, la sous-unité FRA-1 (*Fos-related antigen 1*) du FT AP-1 (*Activator protein 1*) interagit principalement avec α2, alors que le FT TCF4 (*Transcription factor 4*) est surtout recruté à la chromatine par α8. De façon générale, les isoformes contenant la région AF-1 ont été caractérisées comme de plus fortes activatrices de la transcription. L'isoforme α1 active plus fortement l'expression de l'α1-antitrypsine, du RN CAR (*Constitutive androstane receptor*) (NR1I3) et de l'apolipoprotéine CIII comparativement à α7 (Nakhei *et al.*, 1998 ; Pascussi *et al.*, 2007 ; Torres-Padilla *et al.*, 2001). Certains gènes, comme ceux encodant la transthyrétine et l'α-fétoprotéine, sont cependant transactivés préférentiellement par α7 (Torres-Padilla *et al.*, 2001).

Au niveau du domaine F, les isoformes α1 et α2 sont identiques à l'exception d'une insertion additionnelle de 10 acides aminés chez α2. Ce domaine est considéré comme répresseur, puisqu'il inhibe l'interaction de α1 avec NCoA-1 et NCoA-2. La présence d'une séquence supplémentaire dans le domaine F de α2 réduit cet effet répresseur, menant à une meilleure capacité de transactivation pour cette isoforme (Sladek *et al.*, 1999). L'isoforme α3 possède un domaine F beaucoup plus court et différent au niveau de sa

séquence. Dans les cellules Cos7 et Ins-3, $\alpha 3$ active plus fortement l'expression à partir du promoteur *Hnf1a* de la souris comparativement à $\alpha 1$. Ce profil différentiel est identique pour les isoformes $\alpha 9$ et $\alpha 7$, dans l'ordre, mais les isoformes P2 activent plus faiblement la transcription dans ce contexte (Huang *et al.*, 2008).

Très peu d'information est disponible concernant l'activité transcriptionnelle des isoformes non canoniques. Comme mentionné dans la section précédente, les études ayant porté sur l'isoforme $\alpha 4$ utilisaient une séquence qui s'est révélée être fausse, et ne seront donc pas détaillées ici. L'isoforme $\alpha 11$ n'a jamais été étudiée fonctionnellement. Les isoformes $\alpha 10$ et $\alpha 12$ ont quant à elles été montrées pour pouvoir activer la transcription à partir du promoteur *Hnf1a* murin, mais à des niveaux beaucoup plus faibles que les isoformes $\alpha 1$ et $\alpha 3$ (Huang *et al.*, 2009).

5. Hypothèse et objectifs du projet de recherche

Les différentes études portant sur les différences fonctionnelles des isoformes de HNF4 α sont pour l'instant demeurées centrées sur la comparaison des classes P1 et P2. Des rôles distincts ont été associés à ces classes d'isoformes dans différents tissus, incluant le foie, le pancréas et l'intestin. Le domaine de liaison à l'ADN, qui détermine la spécificité de reconnaissance des gènes cibles de HNF4 α , est identique pour toutes les isoformes. Cela suggère que les différences fonctionnelles observées entre les isoformes seraient plutôt dues à des interactions spécifiques pour certains corégulateurs transcriptionnels. Le domaine A/B des isoformes $\alpha 1$, $\alpha 2$ et $\alpha 3$ contient une région activatrice AF-1 alors que celle-ci est absente chez les isoformes $\alpha 7$, $\alpha 8$ et $\alpha 9$. Cette région est connue pour permettre l'interaction de HNF4 α avec plusieurs corégulateurs. Au total, les classes P1 et P2 englobent jusqu'à 12 isoformes potentiellement exprimées, et la plupart d'entre elles ont été caractérisées de manière superficielle en ce qui a trait à leurs interactions et leurs fonctions transcriptionnelles spécifiques. Quatre possibilités de domaines A/B et trois différents domaines F sont retrouvés à travers les 12 isoformes de HNF4 α , et pourraient ainsi coordonner spécifiquement les interactions des isoformes avec certains corégulateurs transcriptionnels.

L'hypothèse derrière ce projet de recherche est donc que des rôles distincts dans la régulation transcriptionnelle existent non seulement entre les classes P1 et P2 de HNF4 α , mais également entre les 12 isoformes composant ces classes. Ces différences fonctionnelles pourront être en grande partie expliquées par la composition des réseaux d'interactions spécifiques à chaque isoforme. Afin de répondre à cette hypothèse de recherche, quatre objectifs ont été établis :

1. Générer et valider des lignées stables exprimant de façon spécifique chacune des 12 isoformes de HNF4 α .
2. Analyser l'impact fonctionnel de l'expression des isoformes sur le transcriptome des lignées stables nouvellement générées.
3. Étudier les partenaires d'interaction spécifiques pour chaque isoforme.
4. Identifier des corégulateurs spécifiques à certaines isoformes pour expliquer les variations dans leurs fonctions transcriptionnelles.

MATÉRIEL ET MÉTHODES

1. Clonage moléculaire

1.1 Clonage des isoformes de HNF4 α

Les 12 isoformes de HNF4 α ont été clonées dans le vecteur donneur pENTR11 (Thermo Fisher Scientific, Waltham, États-Unis) en trois étapes distinctes, consistant à cloner initialement la séquence commune aux 12 isoformes, à insérer de part et d'autre les différentes extrémités N-terminales (domaines A/B), puis les différentes extrémités C-terminales (domaine F). Le système Gateway (Invitrogen, Carlsbad, États-Unis) a ensuite été utilisé dans le but d'obtenir des constructions permettant l'expression des isoformes en fusion avec les étiquettes protéiques GFP et BioID2.

1.1.1 Clonage de la séquence commune dans pENTR11

La séquence commune aux 12 isoformes a été amplifiée par PCR à partir du plasmide pLenti-HNF4 α 2-GFP. Ce plasmide contient la séquence codante optimisée de l'isoforme α 2 de HNF4 α , et provient du laboratoire du Pr François Boudreau (Babeu, 2016) (Annexe D). La séquence commune amplifiée mesure 1 014 paires de bases, couvrant 71-86 % de la séquence complète des différentes isoformes de HNF4 α , et contient les domaines C, D et E de HNF4 α .

Les oligonucléotides utilisés pour l'amplification (Tableau 11, annexe E) ont été obtenus auprès de la compagnie Integrated DNA Technologies (IDT, San Jose, États-Unis). Ceux-ci permettaient l'ajout de sites de restriction *SpeI* et *SfoI* au sein même de la séquence commune, aux extrémités 5' et 3' respectivement, sans changer la composition en acides aminés de la protéine HNF4 α . Parallèlement, ces amorces ont aussi permis l'ajout en amont et en aval de la séquence des sites de restriction *SalI* et *XhoI*, dans l'ordre. La réaction d'amplification a été effectuée avec l'enzyme ADN polymérase iProof High-Fidelity (Bio-Rad, Hercules, États-Unis) selon les recommandations du détaillant (Tableau 12, annexe E). Les réactions ont été effectuées dans le thermocycleur T100 (Bio-Rad, Hercules, États-Unis) suivant les paramètres détaillés dans le tableau 13 (Annexe E).

Le produit PCR a été purifié sur gel d'agarose grâce au *EZ-10 Spin Column DNA Gel Extraction Kit* (Bio Basic, Markham, Canada). Le produit PCR purifié ainsi que le plasmide pENTR11 ont été digérés par les enzymes de restriction *SalI* et *XhoI* (New England Biolabs, Ipswich, États-Unis) dans le tampon de digestion NEBuffer 3.1 (New England Biolabs, Ipswich, États-Unis) pendant 2 heures à 37 °C, puis purifiés sur gel d'agarose. Le plasmide pENTR11 digéré et purifié a ensuite été déphosphorylé par l'enzyme Antarctic Phosphatase (New England Biolabs, Ipswich, États-Unis) pendant 30 minutes à 37 °C, avant de procéder à la ligation entre le produit PCR et pENTR11 suivant un ratio 5 : 1 (insert : vecteur) avec la T4 ADN ligase (New England Biolabs, Ipswich, États-Unis) pendant 2 heures à température pièce. Le plasmide pENTR11-(séquence commune de HNF4a) a été séquencé via la plateforme de séquençage et de génotypage des génomes (Université Laval, Québec, Canada).

1.1.2 Insertion des extrémités N-terminales dans pENTR11

Les quatre différentes extrémités N-terminales des isoformes de HNF4a, cernées des sites de restriction *SalI* (5') et *SpeI* (3'), ont été obtenues auprès de la compagnie Integrated DNA Technologies (IDT, San Jose, États-Unis) (Annexe D). Les séquences correspondant aux extrémités N-terminales des groupes d'isoformes $\alpha 1/2/3$ et $\alpha 4/5/6$ ont été synthétisées directement sous forme d'ADN double-brin (*gBlocks Gene Fragments*). Les séquences correspondant aux extrémités N-terminales des groupes d'isoformes $\alpha 7/8/9$ et $\alpha 10/11/12$ ont quant à elles été obtenues sous formes d'oligonucléotides simple-brin. Afin d'obtenir une séquence double-brin, les oligonucléotides sens et antisens ont été mélangés à une concentration de 1 μM dans un tampon 5 mM NaCl, 20 mM Tris pH 7,5. L'hybridation des oligonucléotides a été réalisée en plaçant le mélange réactionnel dans un bain à sec chauffant réglé à 98 °C pendant 2 minutes, puis en laissant la réaction retourner à température pièce après la fermeture du bain. Le plasmide pENTR11-(séquence commune de HNF4a) ainsi que les quatre extrémités N-terminales ont été digérés par les enzymes de restriction *SalI* et *SpeI* (New England Biolabs, Ipswich, États-Unis) dans le tampon CutSmart (New England Biolabs, Ipswich, États-Unis) pendant 2 heures à 37 °C, avant d'être purifiés sur gel d'agarose. Une réaction de ligation selon un ratio 10 : 1 (insert : vecteur) avec la T4 ADN ligase pendant 2 heures à température pièce a ensuite été effectuée afin d'insérer chaque extrémité N-terminale en amont de la séquence commune de HNF4a déjà présente

dans pENTR11. Les quatre différents plasmides ainsi obtenus ont été séquencés via la plateforme de séquençage et de génotypage des génomes de l'Université Laval.

1.1.3 Insertion des extrémités C-terminales dans pENTR11

Les trois différentes extrémités C-terminales des isoformes de HNF4a, bordées d'un site de restriction *XhoI* en 3', ont été synthétisées directement sous forme d'ADN double-brin (*gBlocks Gene Fragments*) (IDT, San Jose, États-Unis) (Annexe D). Ces trois séquences ainsi que les quatre plasmides pENTR11 contenant les différentes extrémités N-terminales devant la séquence commune de HNF4a ont été digérés par les enzymes de restriction *SfoI* (seulement pour les plasmides) et *XhoI* (New England Biolabs, Ipswich, États-Unis) dans le tampon CutSmart pendant 2 heures à 37 °C, puis purifiés sur gel. Une réaction de ligation comme décrit à la section 1.1.2 a ensuite été effectuée afin d'insérer chaque extrémité C-terminale en aval de la séquence commune de HNF4a dans pENTR11. L'appariement de trois extrémités C-terminales aux quatre extrémités N-terminales déjà clonées a ainsi généré 12 plasmides pENTR11-HNF4a(1-12), contenant la séquence complète de chacune des 12 isoformes de HNF4a. Ces plasmides ont été séquencés via la plateforme de séquençage et de génotypage des génomes de l'Université Laval.

1.1.4 Réaction LR dans les vecteurs d'expression pcDNA-DEST47, pgLAP5.2 et pgLAP5.2-BioID2-3Xmyc

Les séquences des 12 isoformes clonées dans le vecteur pENTR11 ont été transférées dans les vecteurs d'expression pcDNA-DEST47 (Thermo Fisher Scientific, Waltham, États-Unis), pgLAP5.2 (généré dans le laboratoire du Pr François-Michel Boisvert) et pgLAP5.2-BioID2-3Xmyc (clonage détaillé dans la section 1.2 de ce chapitre) par clonage Gateway, via une réaction LR (Thermo Fisher Scientific, Waltham, États-Unis). Cette réaction permet la recombinaison d'une séquence contenue entre deux sites attL1 et attL2, soit la séquence des isoformes dans pENTR11, dans un vecteur d'expression contenant deux sites attR1 et attR2. La réaction LR consistait à mélanger 37,5 ng du plasmide donneur pENTR11-HNF4a(1-12) (1 µl), 37,5 ng du vecteur d'expression (1 µl) et 0,5 µl du mélange d'enzyme LR Clonase II (Thermo Fisher Scientific, Waltham, États-Unis), suivi d'une incubation de 3 heures à 25 °C. La réaction a ensuite été arrêtée via l'ajout de 0,25 µl de solution de protéinase K (Thermo Fisher Scientific, Waltham, États-Unis) pendant 10 minutes à 37 °C. Les plasmides pcDNA-DEST47-HNF4a(1-12), pgLAP5.2-HNF4a(1-12)

et pgLAP5.2-HNF4 α (1-12)-BioID2-3Xmyc ont été séquencés via la plateforme de séquençage et de génotypage des génomes de l'Université Laval (voir annexe D).

1.2 Clonage du plasmide pgLAP5.2-BioID2-3Xmyc

Le plasmide pgLAP5.2-BioID2-3Xmyc a été cloné à partir du plasmide pgLAP5.2 à l'aide de la stratégie de clonage *Gibson assembly* (New England Biolabs, Ipswich, États-Unis). Des amorces chevauchant le plasmide pgLAP5.2 de part et d'autre du gène encodant l'étiquette protéique GFP ont été obtenues auprès de la compagnie Integrated DNA Technologies (IDT, San Jose, États-Unis) (Tableau 14, annexe E). L'amplification par PCR du plasmide pgLAP5.2 avec ces amorces a été réalisée selon les paramètres précédemment décrits dans la section 1.1.1. Le produit PCR résultant a ensuite été purifié sur gel d'agarose. La séquence nucléotidique encodant l'étiquette protéique BioID2-3Xmyc, cernée de séquences identiques aux amorces décrites dans le tableau 14, a été synthétisée directement sous forme d'ADN double-brin (*gBlocks Gene Fragments*, IDT, San Jose, États-Unis). Une réaction *Gibson assembly* a été réalisée en mélangeant 20 ng de produit PCR pgLAP5.2 purifié, 5 ng du *Gene Fragment* BioID2-3Xmyc et 2,5 μ l du mélange réactionnel *Gibson assembly 2x* (New England Biolabs, Ipswich, États-Unis) dans un volume total de 5 μ l, puis en incubant durant 1 heure à 50 °C. Le plasmide pgLAP5.2-BioID2-3Xmyc résultant a été séquencé via la plateforme de séquençage et de génotypage des génomes de l'Université Laval.

1.3 Clonage du plasmide pgLAP5.2-BioID2-3Xmyc-vide

Le plasmide pgLAP5.2-BioID2-3Xmyc-vide a été cloné afin d'exprimer l'étiquette protéique BioID2-3Xmyc seule comme contrôle dans les expériences de spectrométrie de masse. Pour ce faire, le plasmide pgLAP5.2-BioID2-3Xmyc a été amplifié par PCR de manière à retirer un fragment d'environ 1,8 kb contenant le gène de résistance au chloramphénicol et le gène *ccdB*. Les oligonucléotides utilisés à cette fin ont été obtenus auprès de la compagnie Integrated DNA Technologies (IDT, San Jose, États-Unis) et sont présentés dans le tableau 15 (Annexe E). L'oligonucléotide antisens était phosphorylé à son extrémité 5' de manière à pouvoir effectuer la réaction de ligation décrite plus loin dans cette section. La réaction d'amplification par PCR a été effectuée comme décrit dans la section 1.1.1, en utilisant le tampon GC 5X et en ajoutant 3 % DMSO dans un volume total de 20 μ l. Les

réactions ont été effectuées dans le thermocycleur T100 (Bio-Rad, Hercules, États-Unis) suivant les paramètres détaillés dans le tableau 16 (Annexe E). Le produit PCR a ensuite été digéré avec l'enzyme de restriction *DpnI* (New England Biolabs, Ipswich, États-Unis) en ajoutant 1 µl d'enzyme directement dans 20 µl de réaction PCR pendant 30 minutes à 37 °C. Le produit PCR a été purifié sur gel, puis refermé sur lui-même par ligation avec la T4 ADN ligase (New England Biolabs, Ipswich, États-Unis) pendant 30 minutes à température pièce. Le plasmide pGLAP5.2-BioID2-3Xmyc-*vide* ainsi obtenu a été séquencé via la plateforme de séquençage et de génotypage des génomes de l'Université Laval.

1.4 Clonage des gènes codant pour les protéines ayant un potentiel d'interaction avec les isoformes de HNF4α

Les gènes codant pour 12 partenaires potentiels des différentes isoformes de HNF4α identifiés par spectrométrie de masse quantitative (désignés par « gènes partenaires MS ») ont été clonés dans le but de valider ces interactions. La stratégie de clonage consistait d'abord à extraire l'ARN total de cellules HCT 116 et effectuer une réaction de transcription inverse des ARNm (voir sections 5.1 et 5.2). Les amorces utilisées pour cibler les différents gènes permettaient l'amplification du gène (sans codon stop) cerné de séquences attB1 et attB2 (Tableau 17, annexe E) (IDT, San Jose, États-Unis). Les enzymes ADN polymérases iProof High-Fidelity (Bio-Rad) et TransStart KD Plus (TransGen Biotech, Beijing, Chine) ont été utilisées pour l'amplification selon les recommandations des détaillants (Tableau 18, annexe E). Les réactions de PCR ont été réalisées dans le thermocycleur T100 (Bio-Rad, Hercules, États-Unis) suivant les paramètres détaillés dans le tableau 19 (Annexe E). Les gènes ont été clonés initialement dans le plasmide pDONR221 (Thermo Fisher Scientific, Waltham, États-Unis) par réaction BP (système Gateway). Cette réaction permet la recombinaison d'une séquence contenue entre deux sites attB1 et attB2, soit la séquence des gènes amplifiés par RT-PCR, dans un vecteur contenant deux sites attP1 et attP2. La réaction BP consistait d'abord à concentrer 75 ng de produit PCR par évaporateur centrifuge à 65 °C durant 30 minutes, puis à ajouter 75 ng de plasmide pDONR221 (1 µl), 3 µl d'eau stérile et 1 µl du mélange d'enzyme BP Clonase II (Thermo Fisher Scientific, Waltham, États-Unis), suivi d'une incubation de 16 à 18 heures à 25 °C. La réaction a été arrêtée par l'ajout de 0,5 µl de solution de protéinase K pendant 10 minutes à 37 °C. Les plasmides pDONR221-(gènes partenaires MS) ont été séquencés via la plateforme de séquençage et

de géotypage des génomes de l'Université Laval. Les séquences des gènes ont ensuite été clonées dans le vecteur d'expression pDEST-pcDNA5-FLAG C-term, qui a été fourni par le laboratoire du Pr Roucou (Université de Sherbrooke), et généré dans le laboratoire de la Pre Gingras (University of Toronto). Une réaction LR (système Gateway, Thermo Fisher Scientific, Waltham, États-Unis) a été réalisée à cet effet, comme décrit dans la section 1.1.4, le plasmide donneur étant pDONR221-(gènes partenaires MS) dans ce cas. Les plasmides pDEST-pcDNA5-(gènes partenaires MS)-FLAG C-term ont été séquencés via la plateforme de séquençage et de géotypage des génomes de l'Université Laval.

2. Culture cellulaire

Les lignées cellulaires utilisées au cours de ce projet ont été cultivées dans des boîtes de Pétri de polystyrène à 37 °C dans une atmosphère contrôlée à 5 % CO₂, et maintenues à sous-confluence. Les passages cellulaires ont été réalisés en incubant à température pièce les cellules dans une solution de 0,25 % trypsine / 2,21 mM EDTA (Wisent, Saint-Jean-Baptiste, Canada) jusqu'à leur détachement complet, puis en procédant à l'ensemencement d'une dilution appropriée de cellules.

2.1 Lignées cellulaires HCT 116, HCT 116 Flp-In T-REx, HCT 116 HNF4α(1-12)-GFP, HCT 116 HNF4α(1-12)-BioID2-3Xmyc et HCT 116 BioID2-3Xmyc-vide

Les cellules HCT 116 (CCL-247, American Type Culture Collection (ATCC), Manassas, États-Unis) sont des cellules isolées d'un carcinome colorectal humain. La lignée cellulaire HCT 116 Flp-In T-REx a été générée à l'aide du système Flp-In T-REx (Thermo Fisher Scientific, Waltham, États-Unis) et provient du laboratoire du Pr François-Michel Boisvert. Celui-ci permet l'intégration d'une séquence à un endroit précis du génome de ces cellules en une seule copie, l'expression de cette séquence tombant alors sous le contrôle d'un promoteur CMV inductible à la doxycycline. Les lignées stables HCT 116 HNF4α(1-12)-GFP, HCT 116 HNF4α(1-12)-BioID2-3Xmyc et HCT 116 BioID2-3Xmyc-vide ont été générées grâce à ce système (voir section 3.2). La culture de l'ensemble de ces lignées cellulaires a été faite dans du milieu de culture DMEM (*Dulbecco's Modified Eagle's Medium*) (Thermo

Fisher Scientific, Waltham, États-Unis) supplémenté avec 10 % sérum bovin fœtal (FBS - *Fetal Bovine Serum*) de qualité premium (Wisent, Saint-Jean-Baptiste, Canada), 100 U/ml pénicilline et 100 µg/ml streptomycine (Wisent, Saint-Jean-Baptiste, Canada). Les cellules HCT 116 Flp-In T-REx ont également été cultivées en présence des antibiotiques de sélection blasticidine S (5 µg/ml) (Wisent, Saint-Jean-Baptiste, Canada) et zéocine (100 µg/ml) (Thermo Fisher Scientific, Waltham, États-Unis). Les lignées stables HCT 116 HNF4α(1-12)-GFP, HCT 116 HNF4α(1-12)-BioID2-3Xmyc et HCT 116 BioID2-3Xmyc-vide ont quant à elles été maintenues par l'utilisation des antibiotiques de sélection blasticidine S (5 µg/ml) et hygromycine B (100 µg/ml) (Wisent, Saint-Jean-Baptiste, Canada).

2.2 Lignée cellulaire 293T

Les cellules 293T (CRL-3216, ATCC, Manassas, États-Unis) proviennent de cellules de rein embryonnaire humain et sont dérivées de la lignée cellulaire HEK 293 (CRL-1573, ATCC, Manassas, États-Unis). La culture des cellules 293T a été faite dans du milieu de culture DMEM supplémenté avec 10 % FBS de qualité premium, 100 U/ml pénicilline et 100 µg/ml streptomycine.

2.3 Lignées cellulaires Caco-2/15, Capan-2, et Hep G2

Les cellules Caco-2/15 proviennent d'un clone dérivé de la lignée Caco-2 (HTB-37, ATCC, Manassas, États-Unis), provenant elles-mêmes d'un adénocarcinome colorectal humain (Beaulieu et Quaroni, 1991). Les cellules Capan-2 (HTB-80, ATCC, Manassas, États-Unis) sont dérivées d'un adénocarcinome pancréatique humain. La lignée Hep G2 (HB-8065, ATCC, Manassas, États-Unis) a pour origine un carcinome hépatocellulaire humain. La culture de ces lignées cellulaires a été faite dans du milieu de culture DMEM (Wisent, Saint-Jean-Baptiste, Canada) supplémenté avec 10 % FBS de qualité premium, 10 mM HEPES (Wisent, Saint-Jean-Baptiste, Canada) et 2 mM GlutaMAX (Thermo Fisher Scientific, Waltham, États-Unis), 100 U/ml pénicilline et 100 µg/ml streptomycine.

2.4 Lignées cellulaires AsPC-1, COLO 205 et DLD-1

Les cellules AsPC-1 (CRL-1682, ATCC, Manassas, États-Unis) proviennent d'un adénocarcinome pancréatique humain. Les lignées COLO 205 (CCL-222, ATCC, Manassas,

États-Unis) et DLD-1 (CCL-221, ATCC, Manassas, États-Unis) sont dérivées d'adénocarcinomes colorectaux humains de type D et C (classification de Dukes) respectivement. La culture de ces lignées cellulaires a été faite dans du milieu de culture RPMI (*Roswell Park Memorial Institute*) (Wisent, Saint-Jean-Baptiste, Canada) supplémenté avec 10 % FBS de qualité premium, 10 mM HEPES et 2 mM GlutaMAX, 100 U/ml pénicilline et 100 µg/ml streptomycine.

2.5 Lignée cellulaire HT-29

Les cellules HT-29 (HTB-38, ATCC, Manassas, États-Unis) sont des cellules isolées d'un adénocarcinome de côlon humain. La culture des cellules a été faite dans du milieu de culture McCoy's 5A (Wisent, Saint-Jean-Baptiste, Canada) supplémenté avec 10 % FBS de qualité premium, 10 mM HEPES et 2 mM GlutaMAX, 100 U/ml pénicilline et 100 µg/ml streptomycine.

2.6 Lignée cellulaire LoVo

Les cellules LoVo (CCL-229, ATCC, Manassas, États-Unis) sont dérivées de cellules métastatiques retrouvées dans la région supraclaviculaire gauche et provenant à l'origine d'un carcinome colorectal humain. La culture des cellules a été faite dans du milieu de culture F-12K (Wisent, Saint-Jean-Baptiste, Canada) supplémenté avec 10 % FBS de qualité premium, 10 mM HEPES et 2 mM GlutaMAX, 100 U/ml pénicilline et 100 µg/ml streptomycine.

2.7 Lignée cellulaire T84

Les cellules T84 (CCL-248, ATCC, Manassas, États-Unis) sont dérivées de cellules métastatiques retrouvées au niveau d'un poumon et provenant à l'origine d'un carcinome colorectal humain. La culture des cellules a été faite dans du milieu de culture DMEM/F-12 (Wisent, Saint-Jean-Baptiste, Canada) supplémenté avec 10 % FBS de qualité premium, 10 mM HEPES et 2 mM GlutaMAX, 100 U/ml pénicilline et 100 µg/ml streptomycine.

2.8 Culture cellulaire en milieu SILAC

Les lignées cellulaires stables HCT 116 HNF4 α (1-12)-BioID2-3Xmyc et HCT 116 BioID2-3Xmyc-*vide* ont été cultivées dans trois différents milieux SILAC désignés comme léger (R0K0), moyen (R6K4) et lourd (R10K8). Les milieux SILAC contenaient du milieu DMEM + 4,5 g/L glucose, - L-glutamine, - L-arginine, - L-lysine, - pyruvate de sodium (cat. # A1443101, Thermo Fisher Scientific, Waltham, États-Unis) supplémenté avec 10 % FBS dialysé trois fois (Thermo Fisher Scientific, Waltham, États-Unis), 10 mM HEPES, 2 mM GlutaMAX, 100 U/ml pénicilline et 100 μ g/ml streptomycine. À ces milieux ont été ajoutés différents isotopes de L-arginine et de L-lysine à des concentrations finales respectives de 42 μ g/ml et 63,5 μ g/ml afin d'obtenir les milieux légers, moyens et lourds. Le milieu léger contenait la L-arginine R0 (Sigma-Aldrich A6969, Saint-Louis, États-Unis) et la L-lysine K0 (Sigma-Aldrich A8662, Saint-Louis, États-Unis). Le milieu moyen contenait la L-arginine R6 (Cambridge Isotope Laboratories, Inc. CLM-2265, Tewksbury, États-Unis) et la L-lysine K4 (Cambridge Isotope Laboratories, Inc. DLM-2640, Tewksbury, États-Unis). Le milieu lourd contenait la L-arginine R10 (Cambridge Isotope Laboratories, Inc. CNLM-539, Tewksbury, États-Unis) et la L-lysine K8 (Cambridge Isotope Laboratories, Inc. CNLM-291, Tewksbury, États-Unis). Les différents milieux de culture SILAC ont ensuite été filtrés sur unité de filtration Stericup (EMD Millipore, Burlington, États-Unis) avant leur utilisation. Les cellules HCT 116 BioID2-3Xmyc, servant de condition contrôle pour les expériences de spectrométrie de masse, ont été cultivées dans le milieu léger. Les lignées HCT 116 HNF4 α (1-6)-BioID2-3Xmyc ont été cultivées dans le milieu moyen alors que les lignées HCT 116 HNF4 α (7-12)-BioID2-3Xmyc ont été cultivées dans le milieu lourd. La culture en milieu SILAC impliquait un minimum de 5 passages à un ratio 1 : 4 aux 2-3 jours dans des boîtes de Pétri de polystyrène 100 ou 150 mm, à 37 °C dans une atmosphère contrôlée à 5 % CO₂.

3. Transfections

3.1 Transfections transitoires

Les transfections transitoires ont été réalisées avec les cellules 293T, HCT 116 et HCT 116 Flp-In T-REx, lorsque celles-ci avaient atteint une confluence entre 50 et 80 %. La lipofectamine LTX (Thermo Fisher Scientific, Waltham, États-Unis) a été utilisée comme agent de transfection. Les quantités totales d'ADN plasmidique transfecté étaient de 800 ng, 2,5 µg et 5 µg pour des plaques de 24 puits, de 6 puits et des boîtes de Pétri de 60 mm, respectivement. Un premier mélange a d'abord été réalisé entre l'ADN plasmidique, un volume de *Plus Reagent* (Thermo Fisher Scientific, Waltham, États-Unis) en µl équivalent à la quantité d'ADN en µg et du milieu de culture réduit en sérum Opti-MEM (Thermo Fisher Scientific, Waltham, États-Unis) afin de compléter la réaction à un volume de 75 µl (24 puits), 150 µl (6 puits) ou 300 µl (60 mm). Un deuxième mélange a été effectué simultanément entre la lipofectamine LTX, dans un volume en µl équivalent à 1,6 fois la quantité d'ADN plasmidique transfecté en µg, ainsi que du milieu Opti-MEM afin de compléter la réaction au même volume que le premier mélange. Ces deux réactions ont été incubées 5 minutes à température pièce avant d'être mélangées et incubées de 20 à 30 minutes à température pièce. Le mélange réactionnel a ensuite été ajouté aux cellules goutte à goutte de manière à obtenir une distribution uniforme. Les cellules ont été incubées à 37 °C entre 24 et 48 heures avant de procéder à l'extraction de l'ARN ou des protéines, dont les protocoles sont détaillés dans les sections 5 et 6 de ce chapitre.

3.2 Génération de lignées cellulaires stables

Les lignées cellulaires stables HCT 116 HNF4α(1-12)-GFP, HCT 116 HNF4α(1-12)-BioID2-3Xmyc et HCT 116 BioID2-3Xmyc-vide ont été générées grâce au système Flp-In T-REx (Thermo Fisher Scientific, Waltham, États-Unis) en utilisant les plasmides pgLAP5.2-HNF4α(1-12), pgLAP5.2-HNF4α(1-12)-BioID2-3Xmyc (voir section 1.2) et pgLAP5.2-BioID2-3Xmyc-vide (voir section 1.3) respectivement.

Les transfections pour les lignées stables ont été réalisées avec les cellules HCT 116 Flp-In T-Rex, lorsque celles-ci avaient atteint une confluence d'approximativement 70 % dans une boîte de Pétri de 60 mm. La lipofectamine LTX (Thermo Fisher Scientific, Waltham,

États-Unis) a été utilisée comme agent de transfection. Une quantité de 500 ng d'ADN plasmidique a été mélangée à 4,5 µg du vecteur d'expression de la Flp-Recombinase pOG44 (Thermo Fisher Scientific, Waltham, États-Unis), 5 µl de *Plus Reagent* (Thermo Fisher Scientific, Waltham, États-Unis) et du milieu Opti-MEM (Thermo Fisher Scientific, Waltham, États-Unis) afin de compléter la réaction à un volume de 300 µl. Un deuxième mélange a été effectué simultanément entre 8 µl de lipofectamine LTX ainsi que 292 µl de milieu Opti-MEM. Ces deux réactions ont été incubées 5 minutes à température pièce avant d'être mélangées et incubées 30 minutes à température pièce. Entretemps, les cellules HCT 116 ont été lavées une fois avec du PBS 1X, puis 2 ml de milieu de culture DMEM (Thermo Fisher Scientific, Waltham, États-Unis) contenant 10 % FBS (Wisent, Saint-Jean-Baptiste, Canada) a été ajouté aux boîtes de Pétri. Le mélange réactionnel a ensuite été ajouté aux cellules goutte à goutte de manière à obtenir une distribution uniforme. Les cellules ont été incubées à 37 °C pendant 24 heures avant de les ensemercer dans une boîte de Pétri 100 mm. La sélection a débuté 24 heures plus tard, en ajoutant au milieu les antibiotiques blasticidine S (5 µg/ml) et hygromycine B (100 µg/ml). La population sélectionnée a été maintenue dans le temps par l'utilisation continue de ces antibiotiques, et l'induction des constructions a été réalisée via l'ajout de 2,5 µg/ml doxycycline (Clontech Laboratories, Mountain View, États-Unis) dans le milieu au moment souhaité.

4. Immunofluorescence

Les essais d'immunofluorescence ont été réalisés sur les lignées cellulaires stables HCT 116 HNF4α(1-12)-GFP et HCT 116 HNF4α(1-12)-BioID2-3Xmyc. Les cellules ont été ensemençées à une confluence de 25 % sur des lamelles de verre 22 x 22 mm (# 2) (Thermo Fisher Scientific, Waltham, États-Unis) dans des plaques 6 puits. Les cellules ont été incubées 48 heures en présence de 2,5 µg/ml doxycycline (Clontech Laboratories, Mountain View, États-Unis), pour atteindre une confluence approximative de 50 à 70 %. Les cellules ont été fixées avec une solution de 4 % paraformaldéhyde dans du PBS 1X pendant 10 minutes à température pièce, puis perméabilisées dans une solution 0,1 % Triton X-100 - PBS 1X pendant 5 minutes à température pièce. Les cellules ont ensuite été bloquées dans une solution 2 % BSA - PBS 1X pendant 30 minutes à température pièce.

L'incubation avec l'anticorps primaire, dilué dans la solution de blocage (Tableau 3), s'est déroulée pendant la nuit à 4 °C dans une chambre noire humide. Les lamelles ont été lavées à cinq reprises dans du PBS 1X avant d'être incubées avec l'anticorps secondaire pendant 45 minutes à température pièce. Les lamelles ont par la suite été incubées dans une solution de PBS 1X contenant 1 µg/µl DAPI (Thermo Fisher Scientific, Waltham, États-Unis) pendant 5 minutes à température pièce. Les lamelles ont été montées sur des lames à l'aide du liquide de montage Immu-Mount (Thermo Fisher Scientific, Waltham, États-Unis). Les images ont été prises à l'aide d'un microscope à épifluorescence Leica DFC300 FX (Leica, Wetzlar, Allemagne) à un grossissement de 400X.

Tableau 3 : Liste des anticorps utilisés pour les essais d'immunofluorescence

Anticorps	Type	Source	Dilution
GFP	Souris monoclonal, primaire	Roche Applied Science, Penzberg, Allemagne - #11814460001	1 : 500
c-Myc	Souris monoclonal, primaire	ATCC, Manassas, États-Unis - CRL-1729 (9E10)	1 : 25
Alexa Fluor 488	Chèvre anti-souris, secondaire	Thermo Fisher Scientific, Waltham, États-Unis - A-11001	1 : 500
Alexa Fluor 594	Chèvre anti-souris, secondaire	Thermo Fisher Scientific, Waltham, États-Unis - A-11020	1 : 500

5. Extraction et analyse des ARN

5.1 Extraction des ARN totaux à partir de cellules en culture

Les ARN totaux des lignées cellulaires HCT 116, Caco-2/15, T84, COLO 205, LoVo, DLD-1, HT-29, HepG2, AsPC-1 et Capan-2 ont été extraits afin d'analyser par PCR le niveau d'expression des isoformes de HNF4α dans ces lignées. Les ARN totaux des lignées cellulaires stables HCT 116 HNF4α(1-12)-GFP ont été extraits afin d'analyser par PCR le niveau de transactivation de plusieurs gènes cibles de HNF4α par ses différentes isoformes. Les lignées cellulaires stables HCT 116 HNF4α(1-12)-GFP ont été induites 48 heures avec 2,5 µg/ml doxycycline (Clontech Laboratories, Mountain View, États-Unis) avant l'extraction

des ARN totaux. Deux méthodes d'extraction ont été utilisées, soit l'extraction phénol-chloroforme et l'extraction sur colonne à membrane de silice.

5.1.1 Extraction phénol-chloroforme

Le réactif QIAzol (Qiagen, Venlo, Pays-Bas) a été utilisé pour l'extraction phénol-chloroforme, selon les recommandations du détaillant. Les cellules, à une confluence approximative de 80 % dans des plaques de 6 puits ou des boîtes de Pétri de 60 mm, ont été lavées une fois avec du PBS 1X, puis incubées pendant 5 minutes avec 1 ml de réactif QIAzol avant d'être récoltées. Le lysat cellulaire a été incubé 3 minutes en présence de 200 µl de chloroforme, puis centrifugé à 12 000 x g pendant 15 minutes à 4 °C. La phase aqueuse a été recueillie, à laquelle a été ajouté un volume de 500 µl d'isopropanol. Après centrifugation à 12 000 x g pendant 10 minutes à 4 °C, le culot a été lavé dans 1 ml d'éthanol 75 %, puis centrifugé à 7 500 x g pendant 5 minutes à 4 °C. Le culot a été séché à l'air pendant approximativement 5 minutes, puis resuspendu dans 30 à 50 µl d'eau stérile. L'ARN a été dosé à l'aide du NanoDrop (Thermo Fisher Scientific, Waltham, États-Unis), analysé sur gel d'agarose 1,2 % et entreposé à -80 °C.

5.1.2 Extraction sur colonne à membrane de silice

La trousse RNeasy (Qiagen, Venlo, Pays-Bas) a été utilisée pour ce type d'extraction. Les cellules, à une confluence approximative de 80 % dans des boîtes de Pétri de 60 mm, ont été lavées une fois avec du PBS 1X. Un volume de 500 µl de tampon RLT a été utilisé pour la lyse des cellules, qui ont ensuite été grattées et récoltées. Le lysat cellulaire a été homogénéisé à l'aide d'une seringue de 1 ml équipée d'une aiguille 21G, en triturant le volume total à cinq reprises. Un volume d'éthanol 70 % a été ajouté au lysat cellulaire, puis le mélange a été déposé sur une colonne RNeasy avant de centrifuger à 8 000 x g pendant 15 secondes. Un volume de 700 µl de solution RW1, puis 500 µl de solution RPE à deux reprises ont été ajoutés à la colonne, en centrifugeant à 8 000 x g pendant 15 secondes entre chaque étape. Une centrifugation additionnelle de 2 minutes a été réalisée à 8 000 x g, puis un volume de 40 µl d'eau stérile a été ajouté à la colonne. L'élution de l'ARN a été faite en centrifugeant à 8 000 x g pendant 1 minute. L'ARN a par la suite été dosé à l'aide du NanoDrop (Thermo Fisher Scientific, Waltham, États-Unis), analysé sur gel d'agarose 1,2 % et conservé à -80 °C.

5.2 Transcription inverse

Les transcrits d'ARNm extraits de cellules en culture ont été convertis en ADNc par transcription inverse, avec l'enzyme transcriptase inverse SuperScript IV-RT (Thermo Fisher Scientific, Waltham, États-Unis). Une quantité de 2 µg d'ARN a été ajoutée dans un volume total de 10 µl en complétant avec de l'eau DEPC. Un mélange contenant 2,4 µl de 0,5 µg/µl oligos poly(dT) (Amersham Biosciences, Little Chalfont, Royaume-Uni) et 0,8 µl de 25 mM dNTPs (Amersham Biosciences, Little Chalfont, Royaume-Uni) a été ajouté à l'ARN, puis chauffé pendant 5 minutes à 75 °C et déposé sur glace pendant 5 minutes additionnelles. Un volume de 10 µl d'un mélange réactionnel de RT, décrit dans le tableau 4, a été ajouté à l'ARN. La réaction a été incubée pendant 1 heure à 50 °C, avant de procéder à l'inactivation de la SuperScript IV-RT en chauffant pendant 5 minutes à 95 °C. Les échantillons d'ADNc ont par la suite été entreposés à -20 °C.

Tableau 4 : Liste des réactifs utilisés dans le mélange réactionnel de transcription inverse

Réactif	Volume (µl)
Tampon SSIV-RT - 5X (Thermo Fisher Scientific, Waltham, États-Unis)	4,0
DTT - 100 mM (Thermo Fisher Scientific, Waltham, États-Unis)	1,0
Eau DEPC	1,03
Inhibiteur de RNase Protector - 40 unités/µl (Roche Applied Science, Penzberg, Allemagne)	0,6
Enzyme SuperScript IV-RT - 200 unités/µl (Thermo Fisher Scientific, Waltham, États-Unis)	0,17

5.3 PCR (*Polymerase Chain Reaction*) semi-quantitative

5.3.1 Expression des isoformes dans différentes lignées cellulaires cancéreuses humaines

L'expression des 12 isoformes de HNF4α dans les lignées cellulaires HCT 116, Caco-2/15, T84, COLO 205, LoVo, DLD-1, HT-29, HepG2, AsPC-1 et Capan-2 a été évaluée par RT-PCR. Les oligonucléotides spécifiques à chaque isoforme ainsi qu'aux gènes de référence ont été obtenus auprès de la compagnie Integrated DNA Technologies (IDT, San Jose, États-Unis) (Tableau 20, annexe E). Les réactifs et les paramètres utilisés pour l'amplification sont décrits dans les tableaux 21 et 22 de l'annexe E, respectivement. L'ADNc provenant des lignées HCT 116 HNF4α(1-12)-GFP a été utilisé en tant que contrôle positif pour chaque isoforme. Les niveaux d'expression des gènes *HPRT* et *PUM1* ont été utilisés en tant que références, avec une amplification de 26 cycles à une température d'hybridation de 60 °C et un temps d'élongation de 20 secondes.

5.3.2 Expression des isoformes dans différents tissus humains du système digestif

L'expression des 12 isoformes de HNF4α dans différents tissus humains du tractus gastro-intestinal a été évaluée par PCR à l'aide de la banque d'ADNc *Human Digestive System MTC Panel* (Clontech Laboratories, Mountain View, États-Unis). Les préparations d'ADNc retrouvées dans cette banque proviennent de la combinaison d'ADNc issus de plusieurs individus caucasiens sains âgés entre 18 et 61 ans, et étaient fournies à une concentration de 1,0 ng/μl dans 50 μl pour chacun des tissus. Un total de 12 tissus du système digestif étaient inclus dans la banque, soit le foie, l'estomac, l'œsophage, le duodénum, le jéjunum, l'iléon, l'iléocécum, le cécum, le côlon ascendant, transverse et descendant, ainsi que le rectum. Les amorces spécifiques à chaque isoforme de HNF4α, décrites précédemment dans la section 5.3.1, ont été utilisées afin d'évaluer par PCR l'expression des isoformes dans chacun des tissus énumérés ci-haut. Les réactifs utilisés pour l'amplification sont les mêmes ayant été décrits précédemment dans la section 5.3.1, l'ADN matrice étant dans ce cas-ci un volume de 3,7 μl d'ADNc. Les réactions PCR ont été réalisées selon les conditions PCR détaillées dans la section précédente, pour une première ronde de 30 cycles. Une aliquote de 10 μl de la réaction a été prélevée, et le reste de la

réaction a été complétée de nouveau à 20 µl aux concentrations réactionnelles initiales pour une deuxième ronde de PCR de 6 cycles additionnels. Les niveaux d'expression des gènes *POLR2A* et *PSMB2* ont été utilisés en tant que références, avec une amplification de 26 cycles. Un produit PCR correspondant à chaque isoforme a été séquencé pour s'assurer de la spécificité d'amplification. Le plasmide pUC19 a été digéré par l'enzyme de restriction *SmaI* (New England Biolabs, Ipswich, États-Unis) dans le tampon CutSmart (New England Biolabs, Ipswich, États-Unis) pendant 1 heure à 25 °C, puis purifié sur gel d'agarose (voir section 1.1.1). Une ligation entre le produit PCR et le plasmide pUC19 digéré a été réalisée dans un ratio 20 : 1 (insert : vecteur) avec la T4 ADN ligase (New England Biolabs, Ipswich, États-Unis) pendant 3 heures à température pièce. Le séquençage a été effectué via la plateforme de séquençage et de génotypage des génomes (Université Laval, Québec, Canada).

5.4 PCR quantitative en temps réel

La transactivation de différents gènes cibles de HNF4α par ses isoformes a été analysée par PCR quantitative en temps réel (RT-qPCR), à l'aide de l'appareil LightCycler 96 (Roche Applied Science, Penzberg, Allemagne). L'ADNc provenant des lignées HCT 116 HNF4α(1-12)-GFP induites pendant 48 heures a été utilisé. Les oligonucléotides utilisés pour l'amplification des gènes testés ont été obtenus auprès de la compagnie Integrated DNA Technologies (IDT, San Jose, États-Unis) (Tableau 23, annexe E). Les réactions ont été réalisées avec le mélange réactionnel *Fast Start SYBR Green Master* (Roche Applied Science, Penzberg, Allemagne) selon les recommandations du détaillant (Tableau 24, annexe E). Le protocole d'amplification utilisé est décrit dans le tableau 25 (Annexe E). L'analyse des courbes d'amplification et de fusion a été effectuée à l'aide du logiciel LightCycler 96 version 1.1.0.1320 (Roche Applied Science, Penzberg, Allemagne). L'expression relative des gènes a été calculée par comparaison au gène de référence *TBP* avec la formule

$$Eff_{\text{cible}}^{(Cq_{\text{Calibrateur}} - Cq_{\text{Échantillon}})} \times Eff_{\text{référence}}^{(Cq_{\text{Échantillon}} - Cq_{\text{Calibrateur}})}.$$

6. Extraction protéique

6.1 Extraction des protéines totales

Les cellules ont été lavées deux fois dans du PBS 1X, puis lysées en ajoutant directement du tampon Laemmli 2X (4 % SDS, 125 mM Tris pH 6,8, 10 % glycérol). Les lysats cellulaires ont été récoltés à l'aide d'un grattoir, puis soniqués avec le *Sonic Dismembrator Model 120* (Thermo Fisher Scientific, Waltham, États-Unis) afin de réduire la viscosité des échantillons. Les paramètres utilisés consistaient à soniquer les extraits totaux à une amplitude de 25 % trois fois pendant 10 secondes, avec des pauses de 10 secondes sur glace. Les échantillons ont ensuite été centrifugés à 9 600 x g pendant 5 minutes afin de se débarrasser des débris cellulaires. La concentration protéique des échantillons a été dosée à l'aide de la trousse *Pierce BCA Protein Assay* (Thermo Fisher Scientific, Waltham, États-Unis) selon les recommandations du détaillant.

6.2 Extraction des protéines nucléaires

L'extraction de protéines nucléaires a été réalisée sur la lignée cellulaire 293T transfectée avec les différents plasmides pcDNA-DEST47-HNF4 α (1-12). Ces extraits nucléaires ont par la suite été utilisés dans les essais de retard sur gel. Les cellules, cultivées à une confluence approximative de 90 % dans des boîtes de Pétri de 60 mm, ont été trypsinisées puis centrifugées à 1 500 x g pendant 5 minutes. Le culot de cellules a été lavé deux fois avec 1 ml de PBS 1X, en centrifugeant de nouveau à chaque lavage. Les cellules ont ensuite été resuspendues dans 1 ml de tampon d'extraction nucléaire A (10 mM HEPES pH 7,9, 10 mM KCl, 1,5 mM MgCl₂, 0,34 M sucrose, 10 % glycérol, 1 mM DTT, et le *cOmplete, EDTA-free Protease Inhibitor Cocktail*, utilisé selon les recommandations du détaillant (Roche Applied Science, Penzberg, Allemagne)). Un volume de 10 μ l de solution 10 % Triton X-100 a été ajouté à la suspension cellulaire pour obtenir une concentration finale de 0,1 % Triton X-100. Les cellules ont été lysées dans ce tampon pendant 8 minutes sur glace, de manière à briser seulement la membrane plasmique tout en conservant l'intégrité de la membrane nucléaire. Les cellules ont été centrifugées à 1 300 x g pendant 5 minutes à 4 °C. Le surnageant, correspondant à la fraction de protéines cytoplasmiques, a été recueilli et conservé à -80 °C. Le culot de noyaux a ensuite été resuspendu dans 150 μ l de tampon d'extraction nucléaire B (1 % Triton X-100, 150 mM NaCl, 20 mM Tris

pH 7,4, 1 mM DTT, et le *cOmplete, EDTA-free Protease Inhibitor Cocktail*, utilisé selon les recommandations du détaillant). La fraction nucléaire a été incubée avec agitation pendant 30 à 60 minutes à 4 °C, avant d'être entreposée à -80 °C. Les fractions nucléaires et cytoplasmiques ont été centrifugées à 20 000 x g pendant 15 minutes afin de se débarrasser des débris cellulaires. Le dosage de ces fractions a été réalisé à l'aide de la trousse *Pierce BCA Protein Assay* (Thermo Fisher Scientific, Waltham, États-Unis) selon les recommandations du détaillant. Des immunobuvardages des protéines HNF4α, histone H3 et GAPDH ont été réalisés afin de valider la qualité de l'extraction nucléaire.

7. Immunobuvardage (Western blot)

7.1 Électrophorèse des extraits totaux de protéines sur gel de polyacrylamide SDS

Les échantillons de protéines ont été chauffés pendant 5 minutes à 100 °C dans du Laemmli 5X (10 % SDS, 312,5 mM Tris pH 6,8, 25 % glycérol) contenant 10 % β-mercaptoéthanol et 0,1 % bleu de bromophénol (ou dans une concentration diluée de Laemmli contenant un pourcentage proportionnel de β-mercaptoéthanol et de bleu de bromophénol). Ces échantillons ont ensuite été dilués à une concentration de Laemmli 1X et chargés sur gel de polyacrylamide dénaturant (SDS). Les gels SDS ont été coulés à l'aide du *SureCast Gel HandCast System* (Invitrogen, Carlsbad, États-Unis), selon un format « mini-gel » (8 x 8 x 0,1 cm). Un gel de séparation (5-12,5 % acrylamide : bis, 0,37 M Tris pH 8,8, 0,1 % SDS) et un gel de concentration (5 % acrylamide : bis, 0,37 M Tris pH 6,8, 0,1 % SDS) ont été coulés dans l'ordre. Une quantité variant entre 7,5 et 40 µg de protéines a été déposée pour chaque échantillon. La migration a été faite dans du tampon d'électrophorèse 1X (25 mM Tris, 192 mM glycine, 0,1 % SDS) à 125 V de 80 à 120 minutes.

7.2 Transfert des protéines

Les protéines ont été transférées sur une membrane de nitrocellulose (GE Healthcare Life Sciences, Marlborough, États-Unis). Le transfert a été effectué à 25 V pendant

10 à 15 minutes à l'aide du système de transfert semi-sec Trans-Blot Turbo (Bio-Rad, Hercules, États-Unis), dans du tampon Bjerrum Schafer-Nielsen 1X (48 mM Tris, 39 mM glycine, 20 % éthanol, 1,3 mM SDS). L'efficacité du transfert a été vérifiée en effectuant une coloration des protéines avec une solution de 0,1 % Ponceau S.

7.3 Détection des protéines

La membrane a été bloquée avec une solution 5 % lait écrémé en poudre dilué dans du tampon *Western Wash* 1X (154 mM NaCl, 10 mM Tris pH 7,4, 0,1 % Tween 20) entre 30 à 60 minutes à température pièce. La membrane a été lavée dans du tampon *Western Wash* 1X trois fois pendant 5 minutes, avant d'être incubée avec l'anticorps primaire pendant une nuit à 4 °C avec agitation (Tableau 5). Les anticorps primaires ont été dilués dans du tampon d'anticorps primaire (1 % BSA, 10 mM Tris pH 7,4, 150 mM NaCl, 0,01 % NaN₃), à l'exception de l'anticorps anti-FLAG qui a été dilué dans du TBS-T - 3 % Blotto (50 mM Tris-HCl, 150 mM NaCl, 0,1 % Tween 20, 3 % lait écrémé). Trois autres lavages de 5 minutes à température pièce ont été effectués le lendemain avec le tampon *Western Wash* 1X. La membrane a été incubée avec l'anticorps secondaire dilué dans du tampon d'anticorps secondaire (2 % BSA, 1 % lait écrémé en poudre, 10 mM Tris pH 7,4, 150 mM NaCl) entre 1 à 2 heures à température pièce avec agitation. La membrane a de nouveau été lavée trois fois pendant 5 minutes à température pièce avec le tampon *Western Wash* 1X. La détection des protéines via les anticorps secondaires marqués de l'enzyme peroxydase de raifort (HRP - *Horseradish Peroxidase*) a été réalisée à l'aide de la trousse *Clarity Western ECL Blotting Substrates* (Bio-Rad, Hercules, États-Unis) selon les recommandations du détaillant. La détection des signaux de chimiluminescence a été effectuée en utilisant le système d'imagerie ChemiDoc MP (Bio-Rad, Hercules, États-Unis) couplé au logiciel Image Lab 6.0 de la même compagnie. Les films autoradiographiques Amersham Hyperfilm ECL (GE Healthcare Life Sciences, Marlborough, États-Unis) ont également été utilisés pour la révélation dans certains cas, et développés dans une chambre noire.

Tableau 5 : Liste des anticorps utilisés pour la détection des protéines par immunobuvardage

Protéine	Poids moléculaire (PM) (kDa)	Anticorps primaire et dilution	Anticorps secondaire et dilution
HNF4α (C-19)	51	Chèvre polyclonal (Santa Cruz Biotechnology, sc-6556) 1 : 2 000	Anti-chèvre-HRP (Santa Cruz Biotechnology, sc-2056) 1 : 2 000
GFP	27	Souris monoclonal (Roche Applied Science, 11814460001) 1 : 1 000	Anti-souris-HRP (Cell Signaling, 7076S) 1 : 10 000
GAPDH	37	Lapin monoclonal (Cell Signaling, 5174S) 1 : 10 000	Anti-lapin-HRP (Cell Signaling, 7074S) 1 : 10 000
Histone H3	17	Lapin polyclonal (Cell Signaling, 9715S) 1 : 3 000	Anti-lapin-HRP (Cell Signaling, 7074S) 1 : 10 000
Myc-tag	≈ 1	Souris monoclonal (ATCC, CRL-1729 (9E10)) 1 : 50	Anti-souris-HRP (Cell Signaling, 7076S) 1 : 10 000
FLAG	≈ 1	Souris monoclonal (Sigma-Aldrich, F1804-200UG) 1 : 1 000	Anti-souris-HRP (Cell Signaling, 7076S) 1 : 10 000
IRF-2BP2	72	Lapin polyclonal (Thermo Fisher Scientific, PA5-55700) 1 : 500	Anti-lapin-HRP (Cell Signaling, 7074S) 1 : 10 000

8. Retard sur gel

8.1 Hybridation des sondes biotinylées

Les oligonucléotides, contenant l'élément de réponse consensus DR1 de HNF4α (DR1 WT) ou une séquence mutée (DR1 MUT) marqués d'une molécule de biotine à l'extrémité 5' (Tableau 6), ont été obtenus auprès de la compagnie Integrated DNA Technologies (IDT, San Jose, États-Unis). L'hybridation des sondes a été réalisée comme détaillé dans la section 1.1.2. Les sondes hybridées ont ensuite été diluées à une concentration de 20 nM pour les réactions de liaison.

Tableau 6 : Liste des oligonucléotides utilisés pour le retard sur gel

Sonde	Oligonucléotide	Séquence (5' → 3') (DR1)
DR1 WT	Sens	Biotine-GGAGATTAGGTCAAAGGTCAGTAGTCG
	Antisens	Biotine-CGACTACTGACCTTTGACCTAATCTCC
DR1 MUT	Sens	Biotine-GGAGATTACCTCTTAGCAGAGTAGTCG
	Antisens	Biotine-CGACTACTCTGCTAAGAGGTAATCTCC

8.2 Réactions de liaison

Les extraits de protéines nucléaires de cellules 293T utilisés dans ces essais ont été préparés selon le protocole d'extraction nucléaire décrit précédemment (voir section 6.2). Les cellules 293T avaient préalablement été transfectées (ou non) avec les plasmides pDEST47-HNF4α(1-12)-GFP exprimant chacune des 12 isoformes de HNF4α pendant 24 heures, selon le protocole de transfection transitoire décrit précédemment (voir section 3.1).

Les réactions de liaison ont été préparées en plusieurs étapes d'incubation à température pièce, dans un volume final de 20 µl. Une première incubation de 10 minutes a été réalisée en mélangeant 6 µl d'extraits nucléaires (6 µg), 2 µl de tampon de liaison 10X (100 mM Tris pH 7,5, 10 mM EDTA, 1 M KCl, 50 % glycérol et 1 mM DTT) et 1 µl de poly[d(I-C)] (1 µg/µl) (cat. # 10108812001, Roche Applied Science, Penzberg, Allemagne). La sonde biotinylée a ensuite été ajoutée dans le mélange à une concentration finale de 1 nM et incubée pendant 25 minutes. Le *supershift* a été réalisé en ajoutant 1,5 µl d'anticorps monoclonal anti-GFP purifié chez la souris (0,4 µg/µl) (cat. # 11814460001, Roche Applied Science, Penzberg, Allemagne) pour une incubation additionnelle de 15 minutes. Un volume de 2 µl de tampon de chargement 6X-EMSA (6X TBE, 30 % glycérol, 0,125 % bleu de bromophénol) a par la suite été ajouté aux réactions avant le chargement sur gel de polyacrylamide non dénaturant.

8.3 Migration sur gel de polyacrylamide non dénaturant

Les gels de polyacrylamide non dénaturants utilisés ont été préparés à l'aide du *SureCast Gel HandCast System* (Invitrogen, Carlsbad, États-Unis), selon un format

« mini-gel » (8 x 8 x 0,1cm) à une concentration de 4 % acrylamide : bis, 10 % glycérol et TBE 0,5X (45 µM Tris, 45 µM acide borique, 1 mM EDTA pH 8,0). Une pré-migration de 60 minutes à 100 V a été faite en utilisant du tampon TBE 0,5X froid comme tampon de migration. Les échantillons ont ensuite été chargés sur le gel et migrés à 100 V pendant 100 minutes.

8.4 Transfert sur membrane de nylon chargée positivement

Les réactions de liaison ont été transférées sur une membrane de nylon chargée positivement Amersham Hybond-N+ (cat. # RPN203B, GE Healthcare Life Sciences, Marlborough, États-Unis). Le transfert a été effectué dans du tampon TBE 0,5X à 200 mA pendant 90 minutes à 4 °C. Un pontage de l'ADN sur la membrane a ensuite été réalisé par rayons ultra-violet grâce à l'appareil UV Stratalinker 2400 (Stratagene, San Diego, États-Unis), équipé de bulbes à lumière UV (254 nm, 5 x 15 W). La fonction *auto-crosslink* a été utilisée, équivalent à une énergie émise de 1 200 µJ (x 100) pendant 45 secondes.

8.5 Détection des sondes biotinylées

Les étapes menant à la détection de la sonde sur la membrane de nylon chargée positivement ont été réalisées en utilisant le *Chemiluminescent Nucleic Acid Detection Module* (Thermo Fisher Scientific, Waltham, États-Unis). Les solutions utilisées dans cette section réfèrent donc aux solutions contenues dans ce module. Toutes les étapes d'incubation se sont déroulées à température pièce avec agitation. La membrane a d'abord été placée dans 15 ml de tampon de blocage pour 15 minutes, puis dans 15 ml de tampon de blocage contenant 50 µl de conjugué streptavidine-peroxydase de raifort stabilisé (dilution 1 : 300) pour 15 minutes. La membrane a par la suite été lavée quatre fois pendant 5 minutes dans 15 ml de tampon de lavage 1X, puis incubée 5 minutes dans 15 ml de tampon d'équilibration du substrat. Pour la révélation, un volume de 1 ml d'un mélange 1 : 1 des solutions de luminol et de peroxyde stable a été déposé sur la membrane après avoir épongé l'excédent de tampon d'équilibration du substrat. La détection des sondes biotinylées a été effectuée en utilisant le système d'imagerie ChemiDoc MP (Bio-Rad, Hercules, États-Unis) et les films autoradiographiques Amersham Hyperfilm ECL (GE Healthcare Life Sciences, Marlborough, États-Unis), comme décrit dans le protocole d'immunobuvardage (voir section 7).

9. Analyse du transcriptome des lignées HCT 116 HNF4a(1-12)-GFP

9.1 Préparation des ARN totaux

Les lignées HCT 116 HNF4a(1-12)-GFP et HCT 116 Flp-In T-Rex ont été ensemencées dans des boîtes de Pétri de 60 mm et incubées pendant 48 heures en présence de 2,5 µg/ml doxycycline. Les antibiotiques de sélection (blastidine, zéocine et hygromycine) ont été retirés du milieu durant cette période afin d'éviter toute interférence différentielle entre les lignées. Les ARN totaux de ces 13 lignées ont été extraits en triplicatas selon le protocole décrit à la section 5.1.2. La concentration ainsi que la qualité des ARN a été évaluée une première fois via la plateforme de RNomique de l'Université de Sherbrooke par NanoDrop (Thermo Fisher Scientific, Waltham, États-Unis) et par 2100 Bioanalyzer (Agilent, Santa Clara, États-Unis). Les échantillons ont ensuite été envoyés au centre d'innovation de l'Université McGill et de Génome Québec. Une autre analyse de l'intégrité des ARN totaux a été réalisée par le centre d'innovation à l'aide du 2100 Bioanalyzer.

9.2 Préparation des librairies et séquençage des ARN

Les librairies ont été générées à partir de 250 ng d'ARN total. L'enrichissement de l'ARNm a été réalisé à l'aide du *NEBNext Poly(A) Magnetic Isolation Module* (New England Biolabs, Ipswich, États-Unis). La synthèse d'ADNc a été effectuée via l'utilisation des modules *NEBNext RNA First Strand Synthesis* et *NEBNext Ultra Directional RNA Second Strand Synthesis* (New England Biolabs, Ipswich, États-Unis). Les dernières étapes pour la préparation des librairies ont été réalisées à l'aide du *NEBNext Ultra II DNA Library Prep Kit for Illumina* (New England Biolabs, Ipswich, États-Unis). Les adaptateurs et les amorces de PCR ont été obtenues auprès de la compagnie New England Biolabs. Les librairies ont été quantifiées en utilisant le *Quant-iT PicoGreen® dsDNA Assay Kit* (Thermo Fisher Scientific, Waltham, États-Unis) et le *Kapa Illumina GA with Revised Primers-SYBR Fast Universal Kit* (Kapa Biosystems, Wilmington, États-Unis). La taille moyenne des fragments d'ARN a été déterminée à l'aide de l'instrument LabChip GX (PerkinElmer, Waltham, États-Unis). Le séquençage des ARN a été réalisé à l'aide de l'appareil NovaSeq 6000 (Illumina, San Diego, États-Unis) suivant un protocole de type S2 PE100.

9.3 Alignement des résultats de séquençage

La qualité des résultats de séquençage a été visualisée à l'aide de l'outil FastQC 0.11.5 (Andrews, 2016). Le tableau 26 (Annexe F) présente le nombre de lectures obtenues pour chaque échantillon. Le tri des données a été réalisé par le logiciel Trimmomatic 0.36 (Bolger *et al.*, 2014). Un seuil minimal de qualité de 25 a été utilisé pour couper les lectures, qui possédaient en général une excellente qualité (moyenne de 36 pour tous les échantillons). Une longueur minimale de 45 paires de bases a également été utilisée comme seuil, et seules les paires ayant conservé leurs lectures dans les deux directions suite à ce tri ont été conservées. Le transcriptome utilisé pour l'alignement des données a été construit à partir du génome humain GRCh38.p12 et des annotations RefSeq (NCBI Homo sapiens Annotation Release 109) (O'Leary *et al.*, 2016). L'alignement au transcriptome et la quantification ont été réalisés grâce à l'outil Kallisto 0.44.0 (Bray *et al.*, 2016). Les résultats de comptes provenant de Kallisto ont été utilisés pour quantifier les transcrits, ainsi que les gènes par addition des transcrits. Le logiciel DESeq2 1.14.1 a été utilisé afin de calculer l'expression différentielle des transcrits et des gènes pour chaque isoforme en comparaison avec l'échantillon contrôle (Love *et al.*, 2014).

10. Spectrométrie de masse

10.1 Culture cellulaire et induction des différentes constructions

La culture en milieu SILAC des lignées cellulaires HCT 116 HNF4 α (1-12)-BioID2-3Xmyc et HCT 116 BioID2-3Xmyc-vide a été réalisée comme décrit dans la section 2.8 de ce chapitre. Les conditions SILAC étaient : HCT 116 BioID2-3Xmyc-vide en milieu léger R0K0, HCT 116 HNF4 α (1-6)-BioID2-3Xmyc en milieu moyen R6K4 et HCT 116 HNF4 α (7-12)-BioID2-3Xmyc en milieu lourd R10K8. Une boîte de Pétri de 150 mm a été utilisée pour chaque condition. Lors du dernier passage en milieu SILAC, de la doxycycline (Clontech Laboratories, Mountain View, États-Unis) a été ajoutée aux cellules à une concentration de 2,5 μ g/ml 48 heures avant l'expérience de *pull-down*, lorsque les cellules étaient à une confluence de 40 à 50 %. Le jour suivant, soit 24 heures avant le *pull-down*, la biotine (Sigma-Aldrich, Saint-Louis, États-Unis) a été ajoutée aux

cellules à une concentration finale de 50 μ M. Lorsque les conditions R0K0, R6K4 et R10K8 n'étaient pas synchronisées au moment du *pull-down*, les cellules étaient lavées trois fois avec du PBS 1X, puis trypsinisées et centrifugées à 1 500 x g pendant 5 minutes à 4 °C. Les culots de cellules étaient de nouveau lavés avec du PBS 1X à deux reprises, puis conservés à -80 °C jusqu'au moment du *pull-down*.

10.2 *Pull-down* des protéines biotinylées

Les cellules en culture ont été lavées trois fois avec du PBS 1X, et les culots de cellules lavées préalablement (s'il y a lieu) ont été décongelés sur glace. Les cellules ont ensuite été lysées dans 1 ml de tampon RIPA (50 mM Tris-HCl, pH 7,5, 150 mM NaCl, 1,5 mM $MgCl_2$, 0,1 % SDS, 1 % IGEPAL CA-630 (Sigma-Aldrich, Saint-Louis, États-Unis), 1 mM PMSF, 0,4 % désoxycholate de sodium, 1 mM DTT, et le *cOmplete, EDTA-free Protease Inhibitor Cocktail*, utilisé selon les recommandations du détaillant (Roche Applied Science, Penzberg, Allemagne)) par boîte de Pétri de 150 mm. Les lysats cellulaires ont été agités par rotation de 20 à 30 minutes à 4 °C, puis soniqués avec le *Sonic Dismembrator Model 120* (Thermo Fisher Scientific, Waltham, États-Unis) à une amplitude de 30 % trois fois pendant 10 à 15 secondes, avec pauses de 5 secondes sur glace. Une aliquote de 40 μ l de chaque échantillon a été prélevée afin de les doser à l'aide de la trousse *Pierce BCA Protein Assay* (Thermo Fisher Scientific, Waltham, États-Unis) selon les recommandations du détaillant. Les échantillons étaient entretemps agités avec rotation pendant 1 heure supplémentaire à 4 °C, après l'ajout de 10 μ l de 0,1 M EGTA (1 mM final). Une combinaison des trois conditions SILAC (léger, moyen et lourd) a été réalisée avant le *pull-down* en mélangeant 3,5 mg d'extrait total de chaque condition et en complétant dans un volume de 3 ml avec le tampon RIPA. Les détails des combinaisons réalisées sont présentés dans le tableau 7. Un volume de 45 μ l de 20 % SDS a été ajouté à chaque échantillon (0,4 % final). Les échantillons ont été agités avec rotation pendant 15 minutes à 4 °C, puis centrifugés à 20 000 x g pendant 20 minutes à 4 °C.

Les billes de Sepharose couplées à la streptavidine haute performance (GE Healthcare Life Sciences, Marlborough, États-Unis) ont été utilisées pour le *pull-down* à raison de 50 μ l par échantillon combiné. Les billes ont été lavées au préalable à trois reprises avec 1 ml de tampon RIPA contenant 0,4 % SDS. Les lavages ont été réalisés en agitant avec rotation

les billes pendant 5 minutes à 4 °C, puis en centrifugeant à 6 000 x g pendant 3 minutes à 4 °C avant de retirer le surnageant. Les billes lavées ont été resuspendues à une concentration de 50 % dans le même tampon et ajoutées aux échantillons dans des tubes de 5 ml. Les échantillons ont été agités par rotation pendant 3 heures à 4 °C. Les billes ont été centrifugées à 6 000 x g pendant 3 minutes à 4 °C, et une aliquote de 100 µl du surnageant a été prélevée. Les billes ont été transférées dans des microtubes *Low Binding* (Sarstedt, Nümbrecht, Allemagne) et lavées selon les paramètres décrits ci-haut. Un premier lavage a été fait avec 1,5 ml de tampon de lavage BioID (2 % SDS, 50 mM Tris-HCl, pH 7,5), puis trois lavages avec 1 ml de tampon RIPA contenant 0,4 % SDS et cinq lavages avec 1 ml de tampon 20 mM ammonium bicarbonate. Une aliquote correspondant à 5 % de la quantité totale de billes a été prélevée. Les billes lavées suite au *pull-down* ont été conservées à -80 °C, de même que toutes les aliquotes recueillies au cours de l'expérience.

Tableau 7 : Combinaisons des lignées cellulaires en conditions SILAC pour le *pull-down* des protéines biotinylées

Nom de l'échantillon	Condition légère (R0K0)	Condition moyenne (R6K4)	Condition lourde (R10K8)
BioID2-1/7	HCT 116 BioID2-3Xmyc-vide	HCT 116 HNF4a1-BioID2-3Xmyc	HCT 116 HNF4a7-BioID2-3Xmyc
BioID2-2/8	HCT 116 BioID2-3Xmyc-vide	HCT 116 HNF4a2-BioID2-3Xmyc	HCT 116 HNF4a8-BioID2-3Xmyc
BioID2-3/9	HCT 116 BioID2-3Xmyc-vide	HCT 116 HNF4a3-BioID2-3Xmyc	HCT 116 HNF4a9-BioID2-3Xmyc
BioID2-4/10	HCT 116 BioID2-3Xmyc-vide	HCT 116 HNF4a4-BioID2-3Xmyc	HCT 116 HNF4a10-BioID2-3Xmyc
BioID2-5/11	HCT 116 BioID2-3Xmyc-vide	HCT 116 HNF4a5-BioID2-3Xmyc	HCT 116 HNF4a11-BioID2-3Xmyc
BioID2-6/12	HCT 116 BioID2-3Xmyc-vide	HCT 116 HNF4a6-BioID2-3Xmyc	HCT 116 HNF4a12-BioID2-3Xmyc

10.3 Réduction, alkylation et digestion des protéines

Tous les tampons utilisés à partir de cette étape ont été préparés avec de l'eau *MS-grade*. Les billes lavées suite au *pull-down* ont été décongelées sur glace. L'étape de réduction des protéines a été réalisée en incubant les billes dans 100 µl de tampon 20 mM ammonium bicarbonate contenant 10 mM DTT (Thermo Fisher Scientific, Waltham, États-Unis) avec agitation (1250 rpm) pendant 30 minutes à 60 °C. L'alkylation des protéines a été réalisée en ajoutant un autre volume de 100 µl de tampon 20 mM ammonium bicarbonate, puis en complétant à 15 mM IAA (Sigma-Aldrich, Saint-Louis, États-Unis) final avant d'agiter pendant 1 heure à température pièce à l'abri de la lumière. L'IAA a ensuite été neutralisé en complétant à 15 mM DTT et en agitant pendant 10 minutes à 37 °C. Les protéines ont été digérées en ajoutant 1 µg de trypsine Pierce *MS-grade* (Thermo Fisher Scientific, Waltham, États-Unis) et en incubant pour la nuit à 37 °C avec agitation.

10.4 Purification des peptides sur colonne C18

La digestion a été arrêtée en ajoutant 1 % acide formique (FA) (Thermo Fisher Scientific, Waltham, États-Unis) au volume total de 200 µl, puis en agitant pendant 5 minutes à température pièce. Les billes ont été centrifugées à 6 000 x g pendant 3 minutes avant de récolter le surnageant et de le transférer dans un nouveau microtube *Low Binding*. Les billes ont été resuspendues dans 200 µl de tampon contenant 60 % acétonitrile (ACN) (Sigma-Aldrich, Saint-Louis, États-Unis) et 0,1 % FA, puis agitées pendant 5 minutes à température pièce. Le surnageant a été récolté et combiné à celui obtenu précédemment. Ces échantillons ont été concentrés par évaporateur centrifuge à 65 °C jusqu'à assèchement complet (environ 2 heures), et resuspendus dans 30 µl de tampon 0,1 % acide trifluoroacétique (TFA) (Sigma-Aldrich, Saint-Louis, États-Unis). Les peptides ont été purifiés avec des embouts de micropipettes ZipTip de format 10 µl contenant une colonne C18 (EMD Millipore, Burlington, États-Unis). Le ZipTip a d'abord été humidifié en aspirant à trois reprises 10 µl de solution 100 % ACN, puis équilibré en aspirant à trois reprises 10 µl de tampon 0,1 % TFA. Chaque échantillon de peptides a été passé sur le ZipTip équilibré en aspirant et en rejetant à 10 reprises un volume de 10 µl de l'échantillon. Cette étape a été effectuée trois fois afin de passer l'échantillon au complet sur la colonne. Le ZipTip a ensuite été lavé en aspirant à trois reprises 10 µl de tampon

0,1 % TFA. L'élution des peptides a été faite dans un nouveau microtube *Low Binding*, en aspirant et en rejetant à 10 reprises un volume de 10 µl de tampon 50 % ACN et 0,1 % FA. Cette étape a été réalisée trois fois pour obtenir un volume final de 30 µl. Les peptides ont été concentrés par évaporateur centrifuge à 65 °C jusqu'à assèchement complet (environ 30 minutes), puis resuspendus dans 25 µl de tampon 1 % FA. Les peptides ont été dosés à l'aide d'un spectrophotomètre NanoDrop (Thermo Fisher Scientific, Waltham, États-Unis) en faisant la lecture de l'absorbance à 205 nm. Les peptides ont ensuite été transférés dans un vial de verre (Thermo Fisher Scientific, Waltham, États-Unis) et conservés à -20 °C jusqu'à leur analyse par spectrométrie de masse.

10.5 Analyse LC-MS/MS des échantillons peptidiques

Les échantillons peptidiques ont été envoyés à la plateforme de protéomique de l'Université de Sherbrooke à des fins d'analyse. Une quantité de 1,5 µg de peptides a été passée sur HPLC Dionex Ultimate 3000 Binary RSLCnano (Thermo Fisher Scientific, Waltham, États-Unis) couplé à un spectromètre de masse Q Exactive Hybrid Quadrupole-Orbitrap (Thermo Fisher Scientific, Waltham, États-Unis) selon les paramètres précédemment établis par le laboratoire Boisvert (Chauvin *et al.*, 2018). En bref, les peptides ont été chargés sur colonne de C18 Acclaim PepMap 100 (Thermo Fisher Scientific, Waltham, États-Unis) selon un débit de 4 µl/min, puis élués selon un gradient de 5-35 % solvant B (90 % acétonitrile, 0,1 % acide formique) pendant 240 minutes avec un débit de 200 nl/min. L'ionisation des peptides a été réalisée à l'aide d'un voltage de 2,0 kV à une température de 40 °C. Les spectres de sondage MS à balayage complet (m/z 350 - 1 600) en mode profil ont été acquis dans l'Orbitrap avec une résolution de 70 000 après une accumulation d'un million d'ions. Les dix ions peptidiques les plus intenses prévisualisés par l'Orbitrap ont été fragmentés par dissociation induite par collision (énergie de collision normalisée à 35 % et résolution de 17 500) après l'accumulation de 50 000 ions. Les temps maximaux étaient de 250 ms pour les analyses complètes et de 60 ms pour les analyses MS/MS. Les états de charge non attribués ainsi que les espèces avec des charges uniques, 7 ou 8 ont été rejetés. La liste d'exclusion dynamique était limitée à 500 entrées avec une période de rétention maximale de 40 secondes et une fenêtre de masse relative de 10 ppm.

10.6 Identification des protéines par analyse MaxQuant

Les fichiers d'analyse MS/MS ont été analysés à l'aide du logiciel MaxQuant version 1.6.2.2 (Cox et Mann, 2008) et de la base de données Uniprot humaine (16/07/2013). Des analyses ont été initialement réalisées séparément pour chaque isoforme afin d'obtenir des ratios d'enrichissement pour la séquence complète de chacune d'entre elles. Une analyse commune a ensuite été effectuée afin d'intégrer tous les fichiers bruts d'analyse MS/MS. Les paramètres par défaut du logiciel MaxQuant ont été utilisés, à l'exception des paramètres suivants : multiplicité de 3 milieux SILAC (R0K0, R6K4 et R10K8), valeurs d'identification « PSM FDR », « Protein FDR » et « Site decoy fraction » de 0,05, décompte de ratio minimal de 1 et sélection de l'option « Re-quantify ». Suite à l'analyse, les résultats ont été triés selon plusieurs paramètres. Les protéines positives pour au moins une des catégories *Reverse*, *Only identified by site* et *Potential contaminants* ont été éliminées, ainsi que les protéines identifiées à partir d'un seul peptide. Les ratios identifiés dans un seul des trois triplicatas pour chaque expérience ont été éliminés. Les ratios identifiés dans deux des trois triplicatas ont été éliminés lorsqu'ils étaient considérés comme trop divergents, c'est-à-dire lorsque l'écart-type était supérieur à la moyenne des deux ratios d'enrichissement. Les valeurs aberrantes pour les ratios mesurés dans les trois triplicatas ont été détectées à l'aide du test de Grubbs selon une valeur $\alpha = 0,05$, puis éliminées. Suite à ce tri, les protéines pour lesquelles aucun ratio n'avait été calculé dans au moins une des expériences ont été retirées. Les analyses d'enrichissement d'annotations GO (*Gene ontology*) ont été effectuées à l'aide de l'outil Panther 13.1 (Mi *et al.*, 2010). Les visualisations « heatmap » ont été créées à l'aide du logiciel Morpheus (Gould, 2018).

11. Validation des interactions par co-immunoprécipitation

11.1 Co-transfection de HNF4 α (1-12)-GFP et des gènes partenaires MS-FLAG

Les cellules 293T,ensemencées dans des plaques 6 puits, ont été co-transfectées avec les constructions pcDNA-DEST47-HNF4 α (1-12) et pDEST-pcDNA5-(gènes partenaires MS)-FLAG C-term selon le protocole décrit dans la section 3.1 de ce chapitre. Un total de

2,5 µg d'ADN a été transfecté, correspondant à 1,25 µg de chaque plasmide. La spécificité d'interaction testée par co-immunoprécipitation (co-IP) dans la section suivante a été validée en co-transfectant 250 ng du vecteur contrôle pLenti6V5GFPtag (laboratoire du Pr Boudreau), 1,25 µg de pDEST-pcDNA5-(gènes partenaires MS)-FLAG C-term et 1 µg du vecteur pcDNA3.1 (+) (Thermo Fisher Scientific, Waltham, États-Unis). Les cellules transfectées ont été incubées pendant 48 heures à 37 °C avant de procéder à l'expérience de co-IP.

11.2 Co-immunoprécipitation

Les cellules ont été lavées une fois avec du PBS 1X, puis lysées dans 500 µl de tampon de lyse douce (1 % triton X-100, 250 mM NaCl, 20 mM Tris pH 7,4, 1 mM PMSF, et le *cOmplete, EDTA-free Protease Inhibitor Cocktail*, utilisé selon les recommandations du détaillant (Roche Applied Science, Penzberg, Allemagne)). Toutes les étapes pour la suite du protocole ont été réalisées sur glace ou à 4 °C. Les lysats ont été agités par rotation pendant 10 minutes, puis soniqués avec le *Sonic Dismembrator Model 120* (Thermo Fisher Scientific, Waltham, États-Unis) à une amplitude de 25 % pendant 7 secondes. Les extraits totaux ont de nouveau été agités par rotation pendant 10 minutes, avant d'être centrifugés 14 000 x g pendant 5 minutes. La concentration protéique des échantillons a été dosée à l'aide de la trousse *Pierce BCA Protein Assay* (Thermo Fisher Scientific, Waltham, États-Unis) selon les recommandations du détaillant. La co-IP a été réalisée en premier lieu contre GFP avec les billes GFP-Trap (Chromotek, Planegg-Martinsried, Allemagne). Afin de valider les résultats, une co-IP dirigée contre FLAG a ensuite été effectuée, en utilisant des billes magnétiques *Magna ChIP Protein G* (EMD Millipore, Burlington, États-Unis) ainsi qu'un anticorps anti-FLAG (cat. # F1804-200UG, Sigma-Aldrich, Saint-Louis, États-Unis). Dans les deux expériences, les billes ont été préalablement lavées à trois reprises avec 1 ml de tampon de lyse, en agitant par rotation à chaque fois pendant 5 minutes, puis en centrifugeant à 6 000 x g pendant 3 minutes. Une quantité variant entre 500 µg et 1,2 mg d'extrait total a été utilisée pour la co-IP, et une aliquote correspondant à 10 % de cette quantité a été conservée à -80 °C pour analyse par immunobuvardage. Pour la co-IP GFP, un volume de 12 µl de billes GFP-Trap a été utilisé directement au lysat, avant d'agiter par rotation pendant 3 heures. Les billes ont ensuite été lavées à trois reprises avec le tampon de lyse, puis à deux reprises avec du PBS 1X. Les billes ont été conservées à -80 °C pour

analyse par immunobuvardage. Pour la co-IP FLAG, un prélavage a d'abord été réalisé en ajoutant 20 μ l de billes magnétiques-protéine G (lavées au préalable comme décrit ci-haut) au lysat et en agitant par rotation pendant 1 heure. Les billes ont ensuite été retirées avant d'incuber le lysat pendant 2 heures avec 4 μ g d'anti-FLAG (1 μ g/ μ l). Un volume de 20 μ l de billes magnétiques-protéine G lavées au préalable a été ajouté au lysat contenant l'anticorps, pour agiter par rotation pendant 1 heure supplémentaire. Les lavages suite à la co-IP FLAG ont été effectués de la même manière que pour la co-IP GFP. Les co-IP ont été analysées par immunobuvardage, avec les anticorps GFP et FLAG décrits dans la section 7.3. Les protéines ont été éluées des billes en faisant chauffer celles-ci pendant 5 minutes à 100 °C dans du tampon Laemmli 1X (2 % SDS, 62,5mM Tris pH 6,8, 5 % glycérol, 0,02 % bleu de bromophénol) contenant 2 % de β -mercaptoéthanol. Une quantité d'extrait total correspondant à 3,3 % de la quantité utilisée pour la co-IP a été déposée sur gel SDS 8 % polyacrylamide, ainsi que 33 % de l'éluat des billes suite à la co-IP. Les autres étapes du protocole ont été réalisées comme décrit dans la section 7 de ce chapitre.

12. Production et infection lentivirales

12.1 Production lentivirale

Les particules lentivirales ont été produites via l'utilisation des vecteurs pLP1, pLP2 et pLP/VSVG (Thermo Fisher Scientific, Waltham, États-Unis). Les cellules 293T ont étéensemencées dans des boîtes de Pétri 100 mm la veille de la transfection de manière à atteindre une confluence de 90 % le jour de la transfection. Les différents vecteurs d'expression de shARN utilisés sont présentés dans le tableau 8 avec leurs séquences cibles. Les vecteurs pGIPZ_shIRF2BP2 et le vecteur shNonTarget (GIPZ Non-silencing Lentiviral shRNA Control) ont été obtenus auprès de la compagnie Dharmacon (Lafayette, États-Unis). Un premier mélange a été réalisé entre les vecteurs pLP1, pLP2, pLP/VSVG et le vecteur shARN (6 μ g d'ADN chacun) dans 1,5 ml total de milieu de culture réduit en sérum Opti-MEM (Thermo Fisher Scientific, Waltham, États-Unis). Un deuxième mélange a été préparé en ajoutant 48 μ l de Lipofectamine 2000 (Thermo Fisher Scientific, Waltham, États-Unis) et en complétant à 1,5 ml avec du milieu Opti-MEM. Ces deux réactions ont été incubées à température pièce pendant 5 minutes, puis mélangées ensemble et incubées pendant

20 minutes. Entretemps, le milieu de culture des cellules 293T a été remplacé par 5 ml de milieu Opti-MEM. Le mélange réactionnel a ensuite été ajouté aux cellules, qui ont été incubées pendant 4 heures à 37 °C. Le milieu de transfection a ensuite été remplacé par 7 ml de milieu de culture DMEM contenant 10 % FBS. Les cellules ont été incubées pendant 48 heures à 37 °C pour permettre la production des particules lentivirales. Le milieu de culture contenant les particules lentivirales a ensuite été récolté, filtré (0,45 µm), aliquoté en volumes de 700 µl et conservé à -80 °C.

Tableau 8 : Vecteurs d'expression des shARN et leur séquence cible

Vecteurs	Séquence cible (5' → 3')
pGIPZ_shIRF2BP2_V2LHS_102009	GCTTCAAGGTCTACAATTT
pGIPZ_shIRF2BP2_V2LHS_102011	CTCTTATGGTCTGTTGTAT
pGIPZ_shIRF2BP2_V2LHS_232930	CCCTTCGAGAGCAAGTTTA
pGIPZ_shIRF2BP2_V3LHS_308785	CTGATCTTAGTAGCAGACA
pGIPZ_shIRF2BP2_V3LHS_400282	TGGTGACAGTTACCGTTCT
pGIPZ_shIRF2BP2_V3LHS_400283	TGCTTCTAGCTGTATTTGT
pGIPZ_NonTarget (#RHS4346)	CTCGCTTGGGCGAGAGTAA

12.2 Infection lentivirale

Les cellules 293T ont étéensemencées dans des plaques de 6 puits pour atteindre une confluence d'environ 50 % le jour de l'infection. Les cellules ont été infectées avec les différents virus contenant les 6 shIRF2BP2 ainsi que le shNonTarget. Un volume de 350 µl de lentivirus contenant 4 µg/ml de polybrène (Sigma-Aldrich, Saint-Louis, États-Unis) a été utilisé pour chaque infection. Les cellules ont été incubées 1 heure en présence de ce mélange, avant de compléter le volume à 2 ml avec du milieu de culture contenant 4 µg/ml de polybrène. Le milieu a été changé après 48 heures avant de débiter la sélection avec la puromycine à une concentration de 1 µg/ml. La sélection s'est déroulée pendant 7 jours, soit le temps nécessaire pour tuer complètement les cellules contrôles non infectées.

RÉSULTATS

1. Générer et valider des lignées stables exprimant de façon spécifique chacune des 12 isoformes de HNF4 α

1.1 Expression endogène des isoformes

1.1.1 Les isoformes sont exprimées différemment dans les tissus gastro-intestinaux humains

Un total de 12 isoformes ont été répertoriées dans la littérature entourant l'épissage alternatif du gène *HNF4A*. L'expression de certaines d'entre elles n'a été rapportée que dans quelques études ($\alpha 4$ - Drewes *et al.*, 1996 ; $\alpha 10$, $\alpha 11$ et $\alpha 12$ - Huang *et al.*, 2009). Les isoformes $\alpha 5$ et $\alpha 6$ sont putatives et leur expression n'a jamais été démontrée de manière spécifique. Afin de valider l'existence de ces 12 isoformes, différents tissus du tractus gastro-intestinal humain ont été sélectionnés, puisque l'intestin est connu pour être le seul organe exprimant à la fois les isoformes P1 et P2 de HNF4 α . Une banque d'ADNc humains provenant d'individus sains a été utilisée afin d'étudier l'expression des isoformes par PCR semi-quantitative (Figure 6). Comme les isoformes varient uniquement via leurs extrémités N- et C-terminales, les seules amorces pouvant être générées de manière spécifique résultaient en l'obtention d'un amplicon trop grand pour une approche de PCR quantitative ($\approx 1\ 200$ pb). Sept isoformes ont été identifiées dans ces tissus (Figure 6). Les isoformes canoniques de HNF4 α (P1 : $\alpha 1$, $\alpha 2$, et $\alpha 3$; P2 : $\alpha 7$, $\alpha 8$ et $\alpha 9$) sont exprimées dans le foie, l'estomac et tous les segments de l'intestin grêle et du côlon. Aucune isoforme n'a été retrouvée dans l'œsophage, qui n'est pas connu pour exprimer HNF4 α (Tanaka *et al.*, 2006). Bien que l'approche soit non quantitative, on peut tout de même observer une prédominance des isoformes P1 dans le foie, ce qui concorde avec ce qui avait déjà été montré dans la littérature (Torres-Padilla *et al.*, 2001). HNF4 $\alpha 4$ est la seule isoforme non canonique identifiée dans cette expérience, et ce uniquement au niveau du foie.

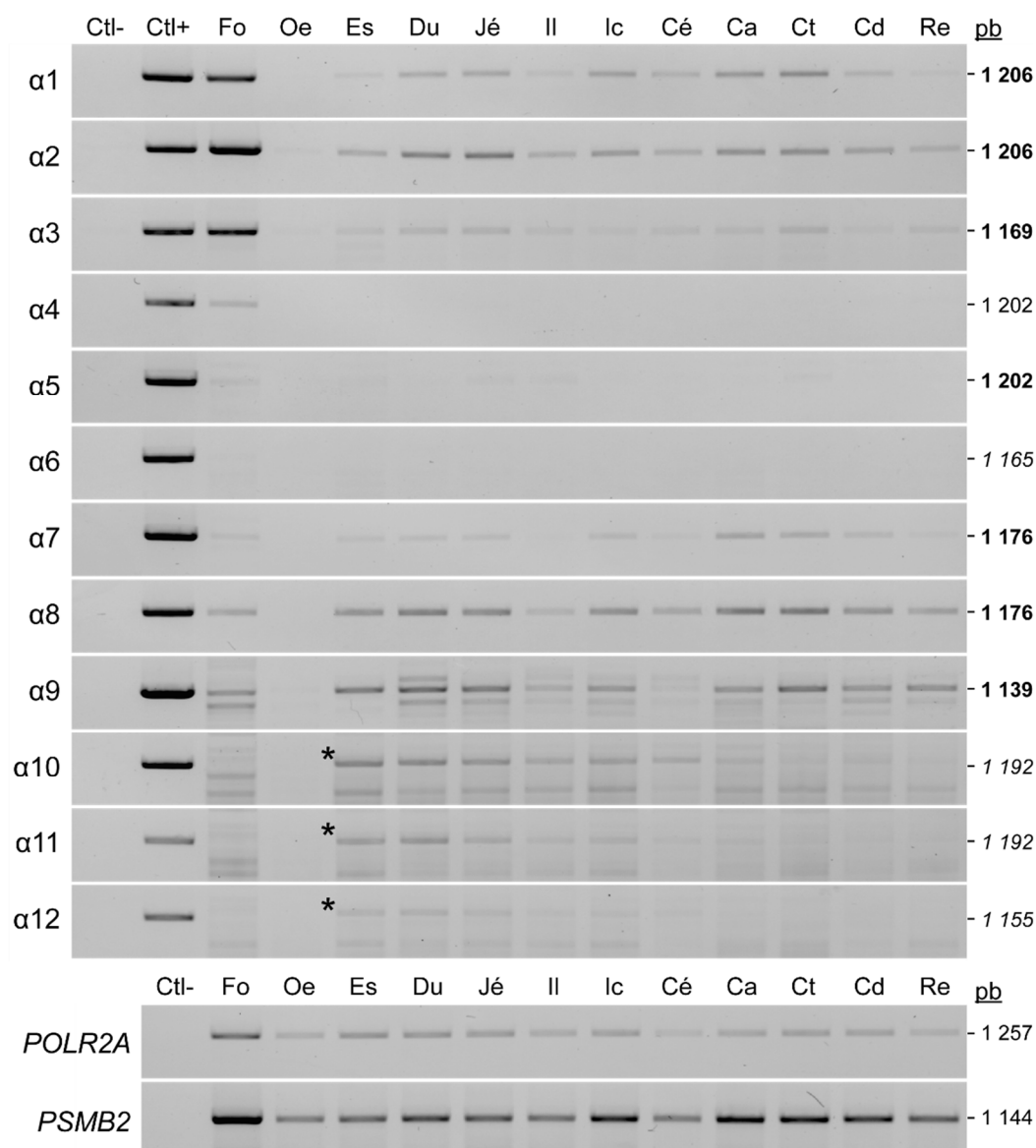


Figure 6 : Les isoformes de HNF4α sont exprimées à des niveaux variables dans les différents tissus du tractus gastro-intestinal humain. L'expression des isoformes a été évaluée par PCR semi-quantitative dans le foie (Fo), l'œsophage (Oe), l'estomac (Es), le duodénum (Du), le jéjunum (Jé), l'iléon (Il), l'iléocécum (Ic), le cécum (Cé), le côlon ascendant (Ca), transverse (Ct) et descendant (Cd), ainsi que le rectum (Re) (n = 1). Les isoformes α5, α6, α10, α11 et α12 n'ont été détectées dans aucun de ces tissus. Les astérisques représentent des bandes déterminées comme non-spécifiques après séquençage. La taille des amplicons est indiquée en nombre de paires de bases (gras = séquencé ; italique = non détecté ; normal = non séquencé). Les gènes *POLR2A* et *PSMB2* ont été utilisés en tant que gènes de référence.

1.1.2 L'expression des isoformes est hautement variable dans plusieurs lignées cellulaires cancéreuses humaines

L'expression des isoformes de HNF4 α est connue pour être modulée dans différents types de cancers (Tanaka *et al.*, 2006). Puisque nous n'avons pas réussi à identifier les 12 isoformes de HNF4 α au niveau des tissus gastro-intestinaux sains, différentes lignées cancéreuses humaines ont été sélectionnées afin de tenter d'observer l'expression de ces isoformes. Comme le pancréas est un organe bien connu pour exprimer HNF4 α , deux lignées d'origine pancréatique, AsPC-1 et Capan-2, ont été incorporées dans cette analyse. L'expression des 12 isoformes a été mesurée dans ces lignées par RT-PCR semi-quantitative (Figure 7), pour les raisons élaborées dans la section précédente. Dans le contexte des lignées cellulaires sélectionnées pour cet essai, aucune isoforme non canonique (α 4-6 et α 10-12) n'a pu être identifiée. Les isoformes canoniques sont quant à elles exprimées dans la plupart de ces lignées, selon un profil similaire à ce qui avait été observé au niveau des tissus sains humains. L'expérience permet également de valider que la lignée HCT 116 Flp-In T-REx utilisée pour la génération de lignées stables dans ce projet n'exprime aucune des 12 isoformes de HNF4 α .

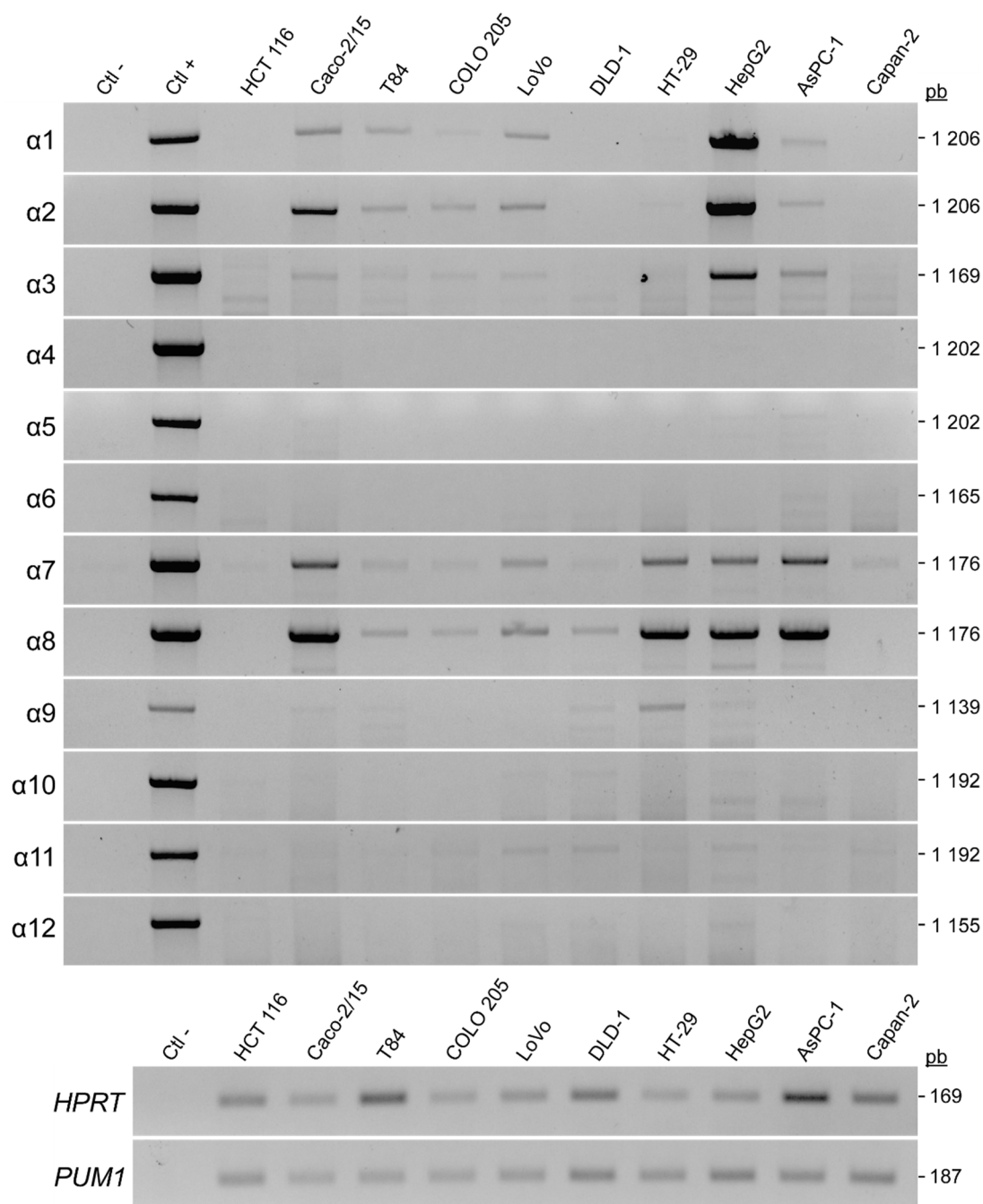


Figure 7 : Les isoformes de HNF4α sont exprimées à des niveaux variables dans différentes lignées cellulaires cancéreuses humaines. L'expression des isoformes a été évaluée par RT-PCR semi-quantitative dans ces lignées (n = 1). Les isoformes non canoniques n'ont pas été détectées dans ces lignées. La taille des amplicons est indiquée en nombre de paires de bases. Les gènes *HPRT* et *PUM1* ont été utilisés en tant que gènes de référence.

1.2 Validation des lignées HCT 116 HNF4α(1-12)-GFP et HCT 116 HNF4α(1-12)-BioID2-3Xmyc

Les lignées cellulaires de cancer colorectal qui montrent une expression endogène de HNF4α expriment toujours plusieurs isoformes de façon concomitante, comme démontré dans la figure 7. Afin d'étudier les fonctions des différentes isoformes de HNF4α de manière spécifique, celles-ci devaient être exprimées dans une lignée cellulaire ne possédant pas d'expression endogène de HNF4α. La lignée cellulaire HCT 116 a été sélectionnée à cet effet, puisque le gène de *HNF4A* est muté dans cette lignée, menant à la perte de son expression (Babeu, 2016 ; Barretina *et al.*, 2012). Cela peut être confirmé par l'expérience de RT-PCR présentée dans la section précédente (Figure 7). Le système Flp-In T-REx a été utilisé afin de générer des lignées stables dans les cellules HCT 116 pour chacune des isoformes de HNF4α. Ce système permet l'incorporation d'un gène à un seul endroit dans le génome de la cellule, sous le contrôle d'un promoteur CMV inducible à la doxycycline. L'avantage de cette stratégie réside dans le fait que les niveaux d'expression des gènes se rapprochent beaucoup plus du niveau d'expression endogène par rapport à une surexpression par transfection, et ces niveaux sont relativement constants pour des gènes de même taille.

1.2.1 Niveaux d'expression protéique des isoformes de HNF4α en fusion avec les étiquettes protéiques GFP et BioID2-3Xmyc

Les différentes isoformes ont dûes être fusionnées à une étiquette protéique GFP, puisqu'aucun anticorps dirigé contre HNF4α n'est capable de reconnaître les 12 isoformes simultanément. D'autres constructions impliquant la fusion des isoformes avec une étiquette protéique BioID2-3xmyc ont été générées pour leur utilisation éventuelle dans une approche BioID couplée à la spectrométrie de masse quantitative. L'étiquette myc a été ajoutée en trois copies à ces constructions, car aucun anticorps n'est disponible pour détecter la protéine BioID2. La fusion de ces étiquettes a été réalisée avec l'extrémité C-terminale des isoformes. Des lignées cellulaires stables exprimant de manière inducible chacune de ces constructions ont été générées dans les cellules HCT 116 Flp-In T-REx. L'expression des constructions a été validée par immunobuvardage, comme montré dans la figure 8. Des variations significatives d'expression ont été observées entre les isoformes, suggérant des

différences dans les niveaux de transcription, de traduction ou encore dans la stabilité des isoformes.

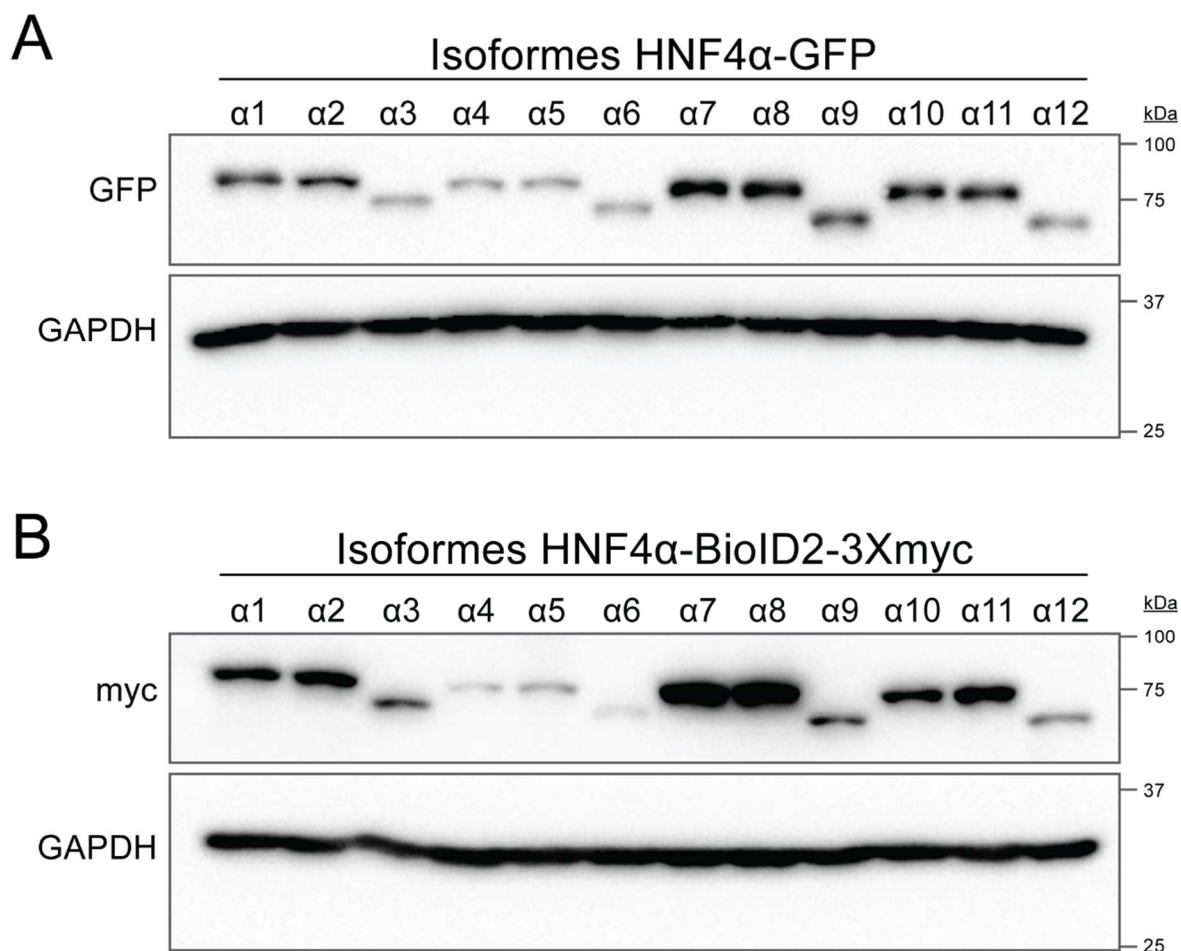


Figure 8 : Les constructions des 12 isoformes de HNF4 α en fusion avec les étiquettes protéiques GFP et BioID2-3Xmyc sont exprimées suite à l'induction des lignées stables dans les cellules HCT 116. Les lignées stables HCT 116 HNF4 α (1-12)-GFP et HCT 116 HNF4 α (1-12)-BioID2-3Xmyc ont été induites pendant 48 heures en présence de 2,5 μ g/ml doxycycline. Des immunobuvardages contre les étiquettes protéiques **(A)** GFP et **(B)** myc ont été réalisés afin de détecter l'expression des différentes protéines de fusion. Les niveaux d'expression de la protéine GAPDH dans ces extraits totaux protéiques ont été utilisés à titre de référence (n = 2).

1.2.2 Les constructions impliquant les différentes isoformes de HNF4α se localisent exclusivement dans le noyau des cellules HCT 116

La localisation nucléaire de HNF4α est essentielle à sa fonction de facteur de transcription. Contrairement à d'autres types de récepteurs nucléaires nécessitant la liaison à leur ligand pour transloquer au noyau, HNF4α est retrouvé constitutivement dans le compartiment nucléaire sous forme d'homodimère. Différents signaux de localisation nucléaire (NLS - *Nuclear Localization Signal*) ont été identifiés dans la séquence de HNF4α, notamment dans son domaine de liaison à l'ADN et sa région charnière (Sun *et al.*, 2007). Plusieurs mécanismes ont été montrés pour réguler la localisation de HNF4α, impliquant notamment diverses modifications post-traductionnelles de la protéine. Afin de valider que les domaines variables des 12 isoformes de HNF4α n'avaient pas d'effet sur leur localisation nucléaire, des essais d'immunofluorescence ont été réalisés sur les lignées stables HCT 116 HNF4α(1-12)-GFP et HCT 116 HNF4α(1-12)-BioID2-3Xmyc. Ces essais confirment que les 12 isoformes de HNF4α se retrouvent exclusivement dans le compartiment nucléaire (Figure 9, en vert), comme le montre le DAPI qui marque les noyaux en bleu.

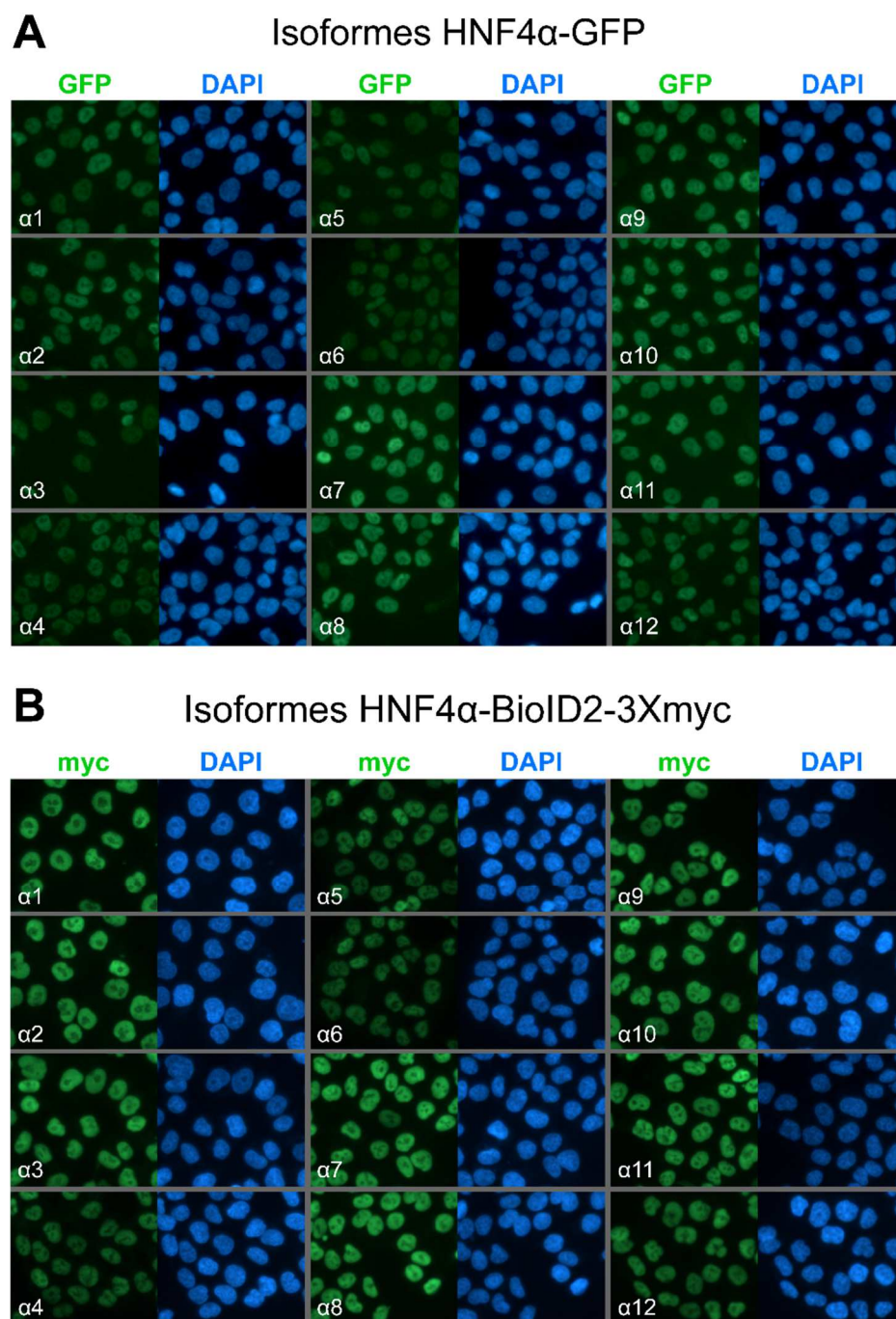


Figure 9 : Les lignées stables HCT 116 exprimant les 12 isoformes de HNF4 α en fusion avec les étiquettes protéiques GFP et BioID2-3Xmyc montrent que celles-ci sont localisées dans le compartiment nucléaire. Les lignées stables HCT 116 HNF4 α (1-12)-GFP et HCT 116 HNF4 α (1-12)-BioID2-3Xmyc ont été induites pendant 48 heures en présence de 2,5 μ g/ml doxycycline. Des essais d'immunofluorescence des étiquettes protéiques **(A)** GFP et **(B)** myc ont été réalisés sur ces cellules afin de détecter l'expression des différentes protéines de fusion (en vert) ($n = 2$). Le marqueur DAPI a été utilisé afin de colorer les noyaux en bleu.

1.2.3 Les isoformes de HNF4a n'ont pas toutes la même capacité de liaison à l'ADN

Après avoir validé l'expression et la localisation des 12 isoformes de HNF4a, nous avons procédé à une première validation fonctionnelle de celles-ci. Comme l'épissage alternatif du gène *HNF4A* touche seulement les domaines N- et C-terminaux, le domaine de liaison à l'ADN est identique d'une isoforme à l'autre. Des essais de retard sur gel ont été réalisés afin de vérifier si les 12 isoformes peuvent lier l'ADN de la même manière (Figure 10). La liaison des isoformes à une sonde contenant l'élément de réponse consensus DR1 de HNF4a (*shift*) a pu être observée pour toutes les isoformes, hormis les isoformes $\alpha 4$, $\alpha 5$ et $\alpha 6$. Un *supershift* réalisé en utilisant un anticorps contre GFP a permis de démontrer que les protéines retardant la sonde sont effectivement les isoformes de HNF4a (fusionnées à GFP). Une sonde possédant une séquence DR1 mutée a été utilisée afin de démontrer la spécificité d'interaction des isoformes vis-à-vis la séquence DR1 consensus.

Cet essai démontre pour la première fois que les isoformes $\alpha 4$, $\alpha 5$ et $\alpha 6$ ne peuvent pas lier l'ADN de la même manière que les autres isoformes de HNF4a, malgré le fait qu'elles possèdent le même domaine de liaison à l'ADN. Cette différence n'est pas attribuable à des niveaux d'expression plus faibles de ces isoformes. Des données non présentées montrent que l'expression des 12 isoformes dans les extraits nucléaires utilisés pour l'essai de retardement sur gel est beaucoup plus similaire d'une isoforme à l'autre comparativement aux extraits totaux présentés précédemment (section 1.2.1).

Ces résultats suggèrent donc que le domaine A/B de ces isoformes puisse réguler négativement leur fonction transcriptionnelle, puisqu'il s'agit de la seule séquence commune pour $\alpha 4$, $\alpha 5$ et $\alpha 6$ et qui n'est pas retrouvée chez les neuf autres isoformes de HNF4a.

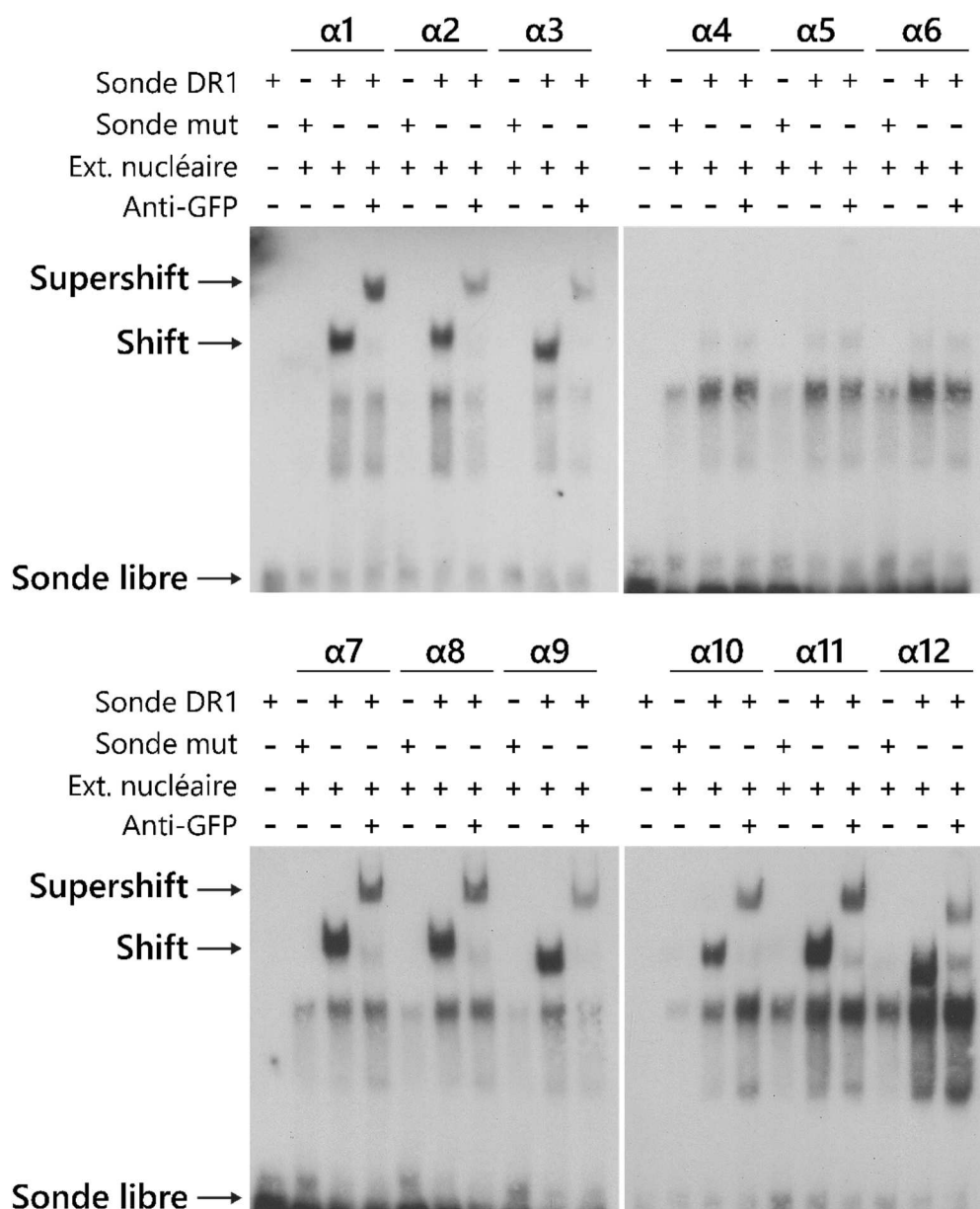


Figure 10 : Les isoformes de HNF4 α ne peuvent pas toutes reconnaître l'élément de réponse consensus DR1 du récepteur nucléaire. Des cellules 293T ont été transfectées pendant 24 heures de manière à surexprimer les différentes constructions HNF4 α (1-12)-GFP. Des extraits nucléaires provenant de ces cellules ont été utilisés afin de réaliser des essais de retard sur gel ($n = 3$). Une sonde d'ADN contenant la séquence consensus DR1 normale ou mutée de HNF4 α biotinylée en 5' a été utilisée. Le *supershift* a été réalisé à l'aide d'un anticorps dirigé contre GFP.

1.2.4 HNF4α active la transcription de ses gènes cibles de façon variable en fonction de ses isoformes

Le récepteur nucléaire HNF4α est connu pour transactiver une multitude de gènes, la majorité d'entre eux étant impliqués dans la différenciation, le métabolisme et le transport de nutriments. Les niveaux d'expression de quatre gènes connus pour être régulés positivement par HNF4α (*APOA1*, *CREB3L3*, *HNF1A* et *VIL1*) impliqués dans ces fonctions ont été mesurés par PCR en temps réel suite à l'expression de chaque isoforme dans les lignées stables HCT 116 HNF4α(1-12)-GFP (Figure 11A). Des différences importantes quant à la transactivation de ces gènes ont été observées entre les isoformes. Le profil de transactivation par les isoformes est similaire d'un gène à l'autre, à quelques exceptions près. Les isoformes α1 et α2 (P1 canoniques) sont les plus fortes activatrices de la transcription des quatre gènes testés, alors que l'isoforme α3 régule positivement *CREB3L3* et *VIL1* à de plus faibles niveaux. Les isoformes α4, α5 et α6 n'ont aucun effet sur les gènes testés, suggérant que la perte de liaison à l'élément de réponse consensus de HNF4α par ces isoformes démontrée dans la section précédente a un impact majeur sur la capacité de transactivation de celles-ci. Les isoformes P2 canoniques (α7, α8) et non canoniques (α10, α11) montrent une capacité de transactivation beaucoup plus faible que les isoformes α1 et α2. Il est intéressant de noter que pour tous les gènes analysés, les variants de chaque sous-groupe partageant un domaine F plus court (α3, α6, α9 et α12) démontrent tous une capacité de transactivation fortement réduite, voire inexistante.

L'analyse des niveaux de transcrits du gène *GFP* pour ces mêmes lignées révèle un profil d'expression des transcrits similaire entre les isoformes de HNF4α (Figure 11B). Comme démontré précédemment dans la section 1.2.1, des différences importantes existent en ce qui concerne l'expression protéique des isoformes. Les résultats présentés à la figure 11B montrent que ces variations dans l'expression protéique ne sont pas dues à des niveaux transcriptionnels différents, et impliqueraient plutôt d'autres mécanismes de régulation post-transcriptionnelle.

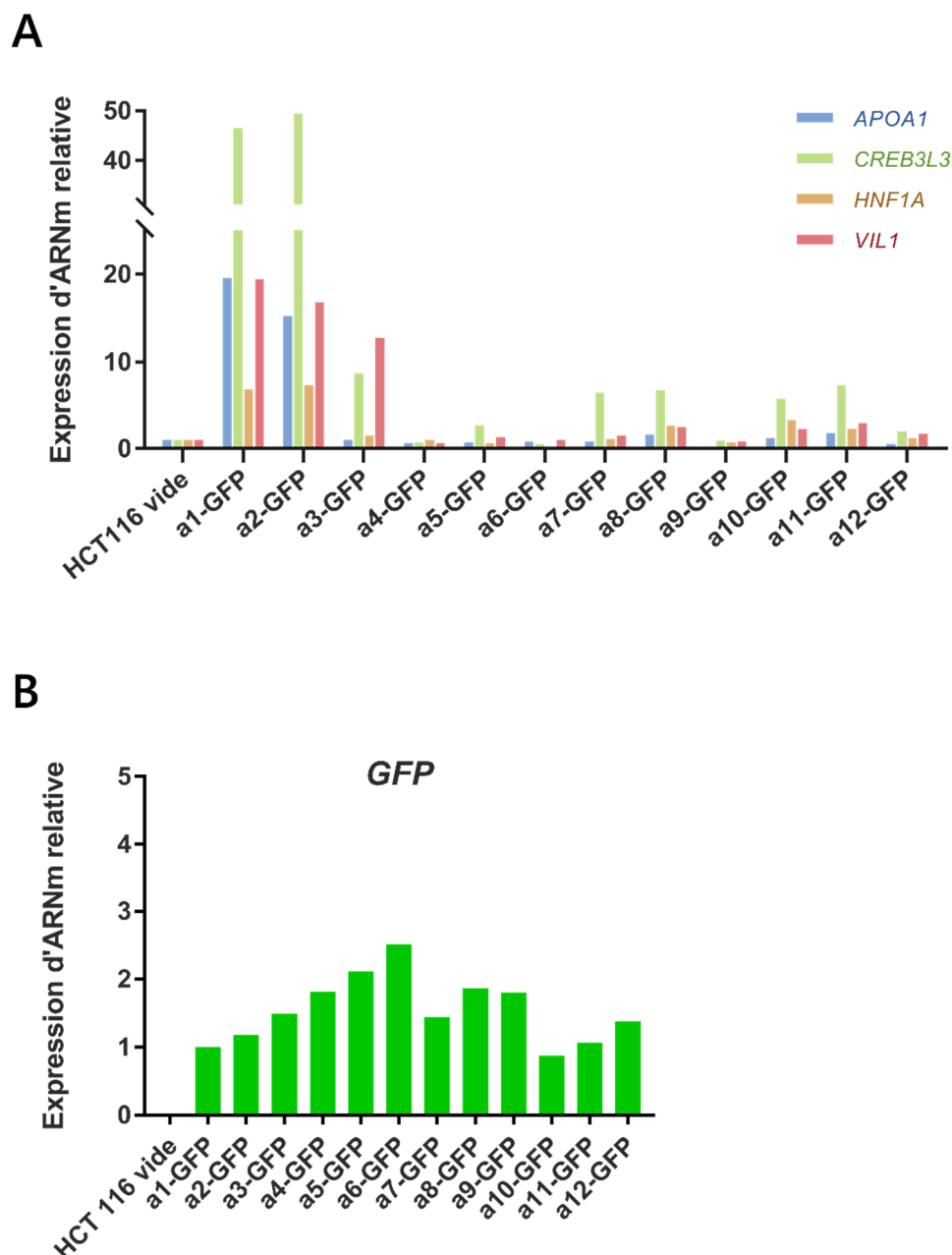


Figure 11 : Des gènes cibles connus de HNF4 α sont transactivés à différents niveaux en fonction des isoformes exprimées. Des réactions de RT-PCR quantitative (qPCR) ont été réalisées sur des échantillons provenant des lignées stables HCT 116 HNF4 α (1-12)-GFP ayant été induites pendant 48 heures en présence de 2,5 μ g/ml doxycycline. **(A)** Les niveaux de transcrits de quatre gènes cibles connus de HNF4 α ont été mesurés, soit *APOA1*, *CREB3L3*, *HNF1A* et *VIL1*. **(B)** Les niveaux de transcrits des isoformes fusionnées à GFP ont été évalués en mesurant l'expression de *GFP*. Les niveaux d'expression ont été relativisés par rapport au gène de référence *TBP* et normalisés par rapport aux cellules HCT 116 vides (n = 1).

2. Analyser l'impact fonctionnel de l'expression des isoformes sur le transcriptome dans le contexte des cellules HCT 116

Afin d'étudier de façon plus approfondie les fonctions transcriptionnelles des 12 isoformes de HNF4 α , nous avons procédé à l'analyse du transcriptome des lignées HCT 116 HNF4 α (1-12)-GFP nouvellement générées. Pour ce faire, les ARN provenant de ces lignées ont été séquencés en triplicatas. Les lectures ont été alignées au transcriptome humain construit à partir des annotations de la base de données RefSeq. L'expression différentielle des transcrits et des gènes a été obtenue en comparant les données spécifiques pour chaque isoforme au transcriptome de la condition contrôle (HCT 116 Flp-In T-REx).

2.1 Les isoformes qui partagent le même domaine A/B sont associées à des profils transcriptomiques plus similaires

Une analyse exploratoire des ensembles de données a été effectuée par PCA (*Principal component analysis*). Ce type d'analyse permet d'observer la variance retrouvée entre les transcrits quantifiés pour l'ensemble des isoformes. La visualisation des deux dimensions par lesquelles s'explique la majeure partie de cette variance est présentée à la figure 12, les composantes principales PC1 et PC2 comptant pour 42,7 % et 13,1 % de cette variance, respectivement. Ce type d'analyse permet en premier lieu de valider la reproductibilité des triplicatas pour chaque expérience, qui sont tous regroupés fortement à l'exception du premier des trois triplicatas pour l'isoforme $\alpha 3$ (symbolisé par A3N1). La divergence de cet échantillon s'explique par une expression plus faible de l'isoforme HNF4 $\alpha 3$ (214 transcrits par million (TPM) contre 460 et 436 TPM dans les deux autres répétitions biologiques). La suite des analyses a donc été effectuée en écartant cet échantillon, et en utilisant la valeur moyenne des décomptes entre les différentes répétitions biologiques.

La condition contrôle est symbolisée par A0 sur la figure 12, et se retrouve complètement à gauche selon la composante PC1 et au centre selon PC2. Plus le point représentant une isoforme est éloigné de cette condition contrôle, surtout selon la dimension horizontale PC1, plus le profil transcriptomique associé à cette isoforme est différent du transcriptome habituel des cellules HCT 116. Ainsi, les isoformes $\alpha 1$ et $\alpha 2$

causent le plus de variance par rapport au contrôle, alors que $\alpha 3$ a un effet plus mitigé. À l'opposé de cette observation, les isoformes $\alpha 4$, $\alpha 5$ et $\alpha 6$ provoquent très peu de divergence. Les isoformes P2 ont quant à elles un effet intermédiaire par rapport à ces deux sous-groupes d'isoformes P1. Il est intéressant de noter que de façon générale, les isoformes semblent se regrouper fortement en fonction de leur domaine A/B, à l'exception d'une divergence plus grande observée entre les isoformes $\alpha 10$, $\alpha 11$ et $\alpha 12$. Cela pourrait signifier que la majorité des différences fonctionnelles entre les isoformes de HNF4 α sont associées à ce domaine plutôt qu'au domaine F. Néanmoins, nous notons un effet au niveau du domaine F plus court, que contiennent les isoformes $\alpha 3$, $\alpha 6$, $\alpha 9$ et $\alpha 12$. Ces isoformes sont toutes plus rapprochées du témoin négatif par rapport aux deux autres isoformes contenant le même domaine A/B.

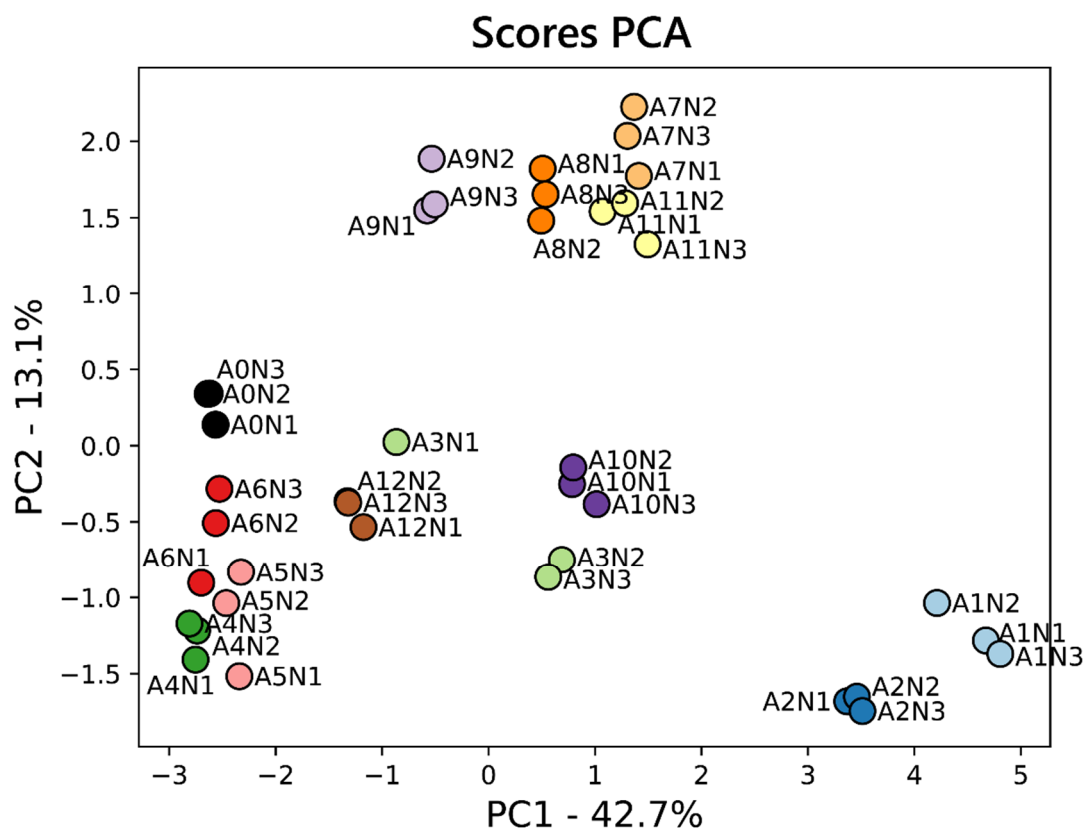


Figure 12 : L'analyse de la variance entre les ensembles de données révèle des profils transcriptomiques plus similaires pour les isoformes ayant un domaine A/B commun.

Analyse PCA des données associées aux 12 isoformes de HNF4 α (A1-A12) et à la condition contrôle (A0). Les triplicatas pour chaque échantillon sont représentés par N1 / N2 / N3 à la suite du numéro d'isoforme. Les deux composantes principales PC1 et PC2 sont utilisées pour la visualisation bidimensionnelle de l'analyse.

2.2 Comparaison du transcriptome des cellules HCT 116 exprimant chacune des 12 isoformes de HNF4a

Environ 35 000 gènes ont pu être quantifiés en fonction de leur modulation pour chaque isoforme par rapport à la condition contrôle. Pour tenter d'expliquer la variance observée dans la section précédente, nous avons choisi de comparer uniquement les gènes dont l'expression était affectée de façon significative. Seuls les gènes modulés positivement ou négativement plus de deux fois par rapport au contrôle ont été considérés, selon un seuil minimal de valeur-p de 0,001. Les résultats peuvent être visualisés de façon globale sous forme de cercles dont la taille est proportionnelle au nombre de gènes modulés (Figure 13). Les niveaux de modulation positive et négative sont aussi présentés sous un aspect de diagramme circulaire. Une représentation « volcano plot » des résultats triés de la même façon permet d'analyser de façon plus détaillée la distribution des gènes retrouvés à travers les différents ensembles de données (Figure 14).

La différence principale pouvant être observée grâce à ces représentations concerne le nombre total de gènes modulés suite à l'expression des 12 isoformes de HNF4a. Les isoformes $\alpha 1$ et $\alpha 2$ sont de loin celles qui ont le plus grand impact sur le transcriptome dans ce contexte. Les isoformes $\alpha 4$, $\alpha 5$ et $\alpha 6$ n'ont quant à elles qu'un effet extrêmement faible sur l'expression des gènes. Un effet intermédiaire est noté pour les isoformes P2 canoniques et non canoniques. De façon générale, il semble que HNF4a soit davantage impliqué dans l'activation de l'expression des gènes qu'à leur répression. Les gènes modulés par les isoformes $\alpha 1$ et $\alpha 2$, qui sont les plus importants régulateurs dans ce contexte, sont majoritairement modulés à la hausse (66 % et 63 %, respectivement). Certaines isoformes comme $\alpha 9$ semblent plutôt être majoritairement responsables d'une répression de la transcription, alors que d'autres comme $\alpha 3$ sont encore plus fortement activatrices. Dans l'ensemble, nous pouvons affirmer qu'il existe des différences majeures dans la fonction transcriptionnelle entre les 12 isoformes de HNF4a. L'analyse plus approfondie des réseaux de gènes régulés spécifiquement par ces isoformes demeure encore à effectuer pour mieux caractériser ces fonctions différentielles.

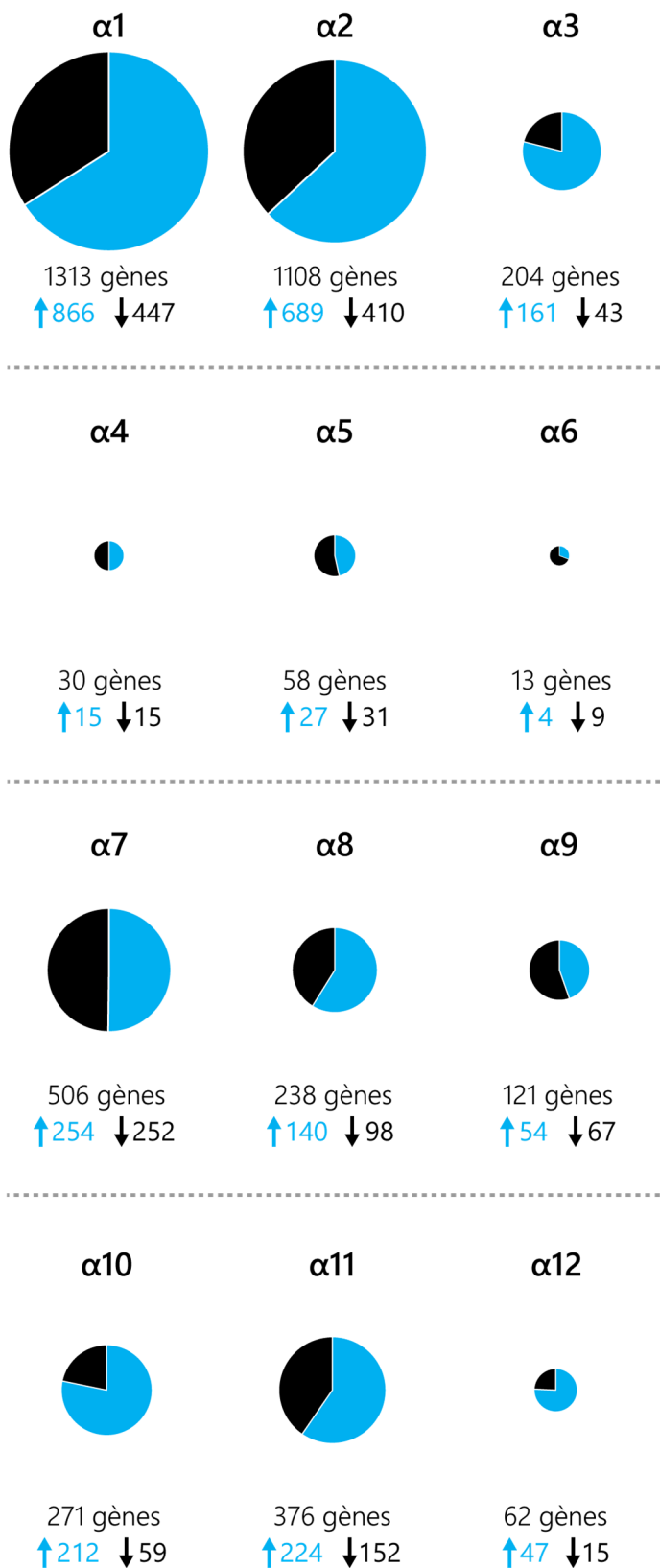


Figure 13 : L'expression des 12 isoformes de HNF4 α provoque des modulations hautement variables du transcriptome des cellules HCT 116. Représentation du nombre de gènes modulés à la hausse et à baisse pour chacune des isoformes de HNF4 α . Un seuil minimal de modulation absolue de 2 combiné à un seuil de valeur-p ajustée $\leq 0,001$ ont été utilisés. La taille des cercles est proportionnelle au nombre total de gènes modulés pour chaque isoforme. La couleur bleue représente la proportion de gènes modulés à la hausse par rapport à la condition contrôle, alors que la portion noire représente les gènes modulés négativement.

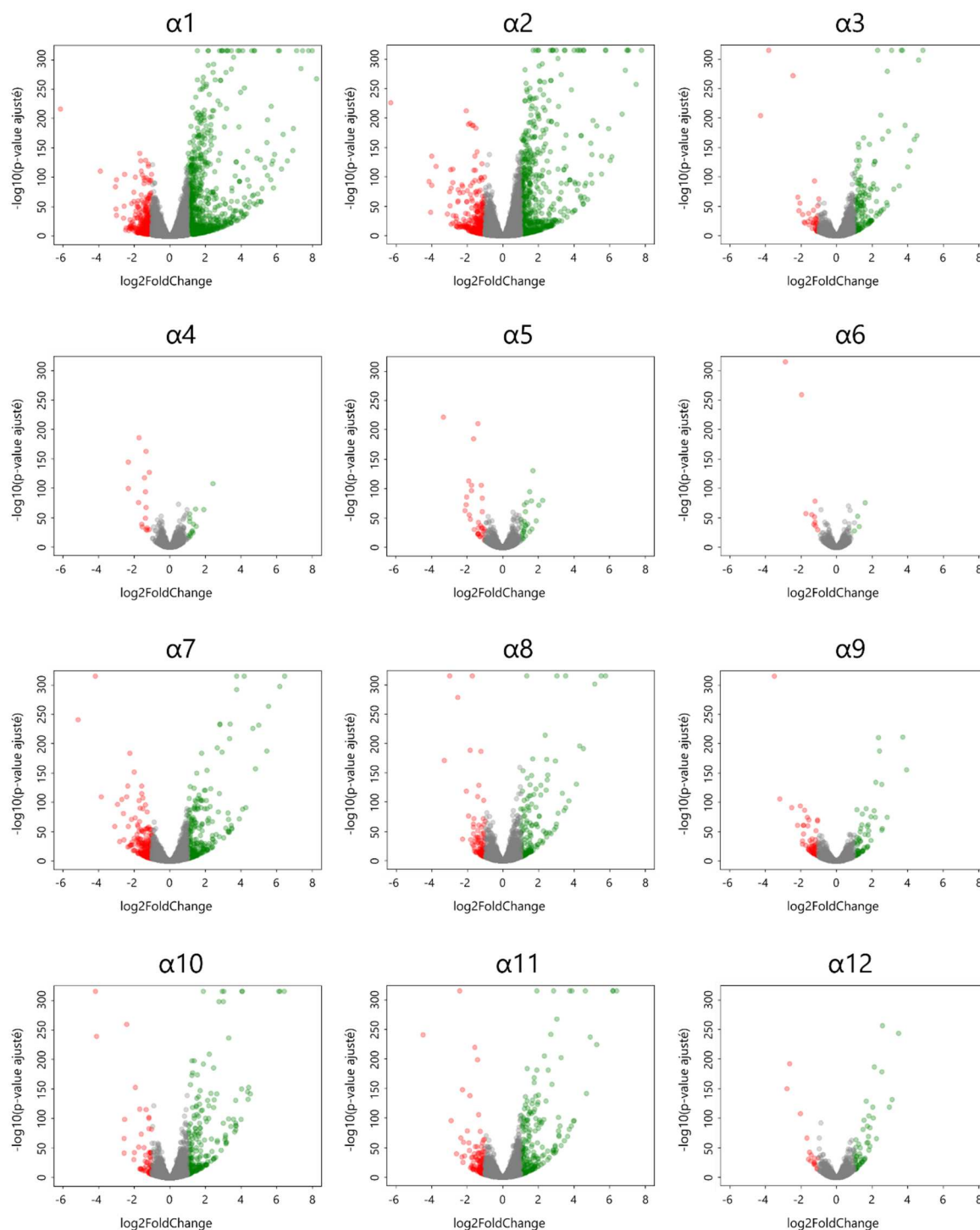


Figure 14 : Représentation « volcano plot » de l'expression différentielle des gènes. Le logarithme en base 10 de la valeur-p ajustée est présenté en fonction du logarithme en base 2 du niveau de modulation (FoldChange). Un seuil minimal de modulation absolue de 2 combiné à un seuil de valeur-p ajustée $\leq 0,001$ ont été utilisés. Les gènes au-dessus de ce seuil sont modulés à la hausse (en vert) ou à la baisse (en rouge).

3. Étudier les partenaires d'interaction spécifiques pour chaque isoforme

3.1 Les 12 isoformes de HNF4α possèdent des réseaux d'interactions distincts, formés en majeure partie de facteurs de transcription et de corégulateurs transcriptionnels

De vastes différences dans la quantité de gènes modulés ont été observées entre les isoformes de HNF4α. L'activité fonctionnelle de HNF4α est régulée par deux caractéristiques reliées à sa structure. Son domaine de liaison à l'ADN lui permet la reconnaissance de séquences régulatrices spécifiques au niveau de ses gènes cibles, alors que différentes régions lui permettent d'interagir avec plusieurs corégulateurs transcriptionnels. Puisque les 12 isoformes possèdent le même DBD, nous avons proposé que l'interaction variable des isoformes avec certains corégulateurs serait l'explication derrière leurs fonctions transcriptionnelles distinctes. Afin de déterminer les réseaux d'interaction protéine-protéine impliquant les isoformes de HNF4α, nous avons utilisé une approche BioID couplée à la spectrométrie de masse quantitative (Figure 15). L'identification de la biotinylation dépendante de la proximité, ou BioID, est une approche utilisant différentes enzymes capables de marquer par la biotine les protéines situées dans l'environnement immédiat d'une protéine d'intérêt. Ce marquage par la biotine permet d'isoler les protéines dans des conditions plus stringentes que par des méthodes d'immunoprécipitation plus classiques, améliorant ainsi la spécificité d'identification subséquente des protéines.

La biotine ligase BirA* fut la première enzyme utilisée pour l'approche BioID (Roux *et al.*, 2012). Une forme plus courte et plus performante développée plus tard, nommée BioID2, a été utilisée dans le cadre de ce projet (Figure 15A) (Kim *et al.*, 2016). Des lignées stables dans la lignée HCT 116 ont été générées à partir de constructions permettant l'expression de chacune des 12 isoformes de HNF4α en fusion avec BioID2. La quantification des protéines a pu être possible grâce à l'utilisation de l'approche SILAC (*Stable isotope labeling with amino acids in cell culture*). Celle-ci consiste en l'utilisation de milieux de culture contenant différents isotopes d'arginine et de lysine, qui sont incorporés dans les protéines des cellules cultivées dans ces milieux. Trois conditions ont ainsi pu être

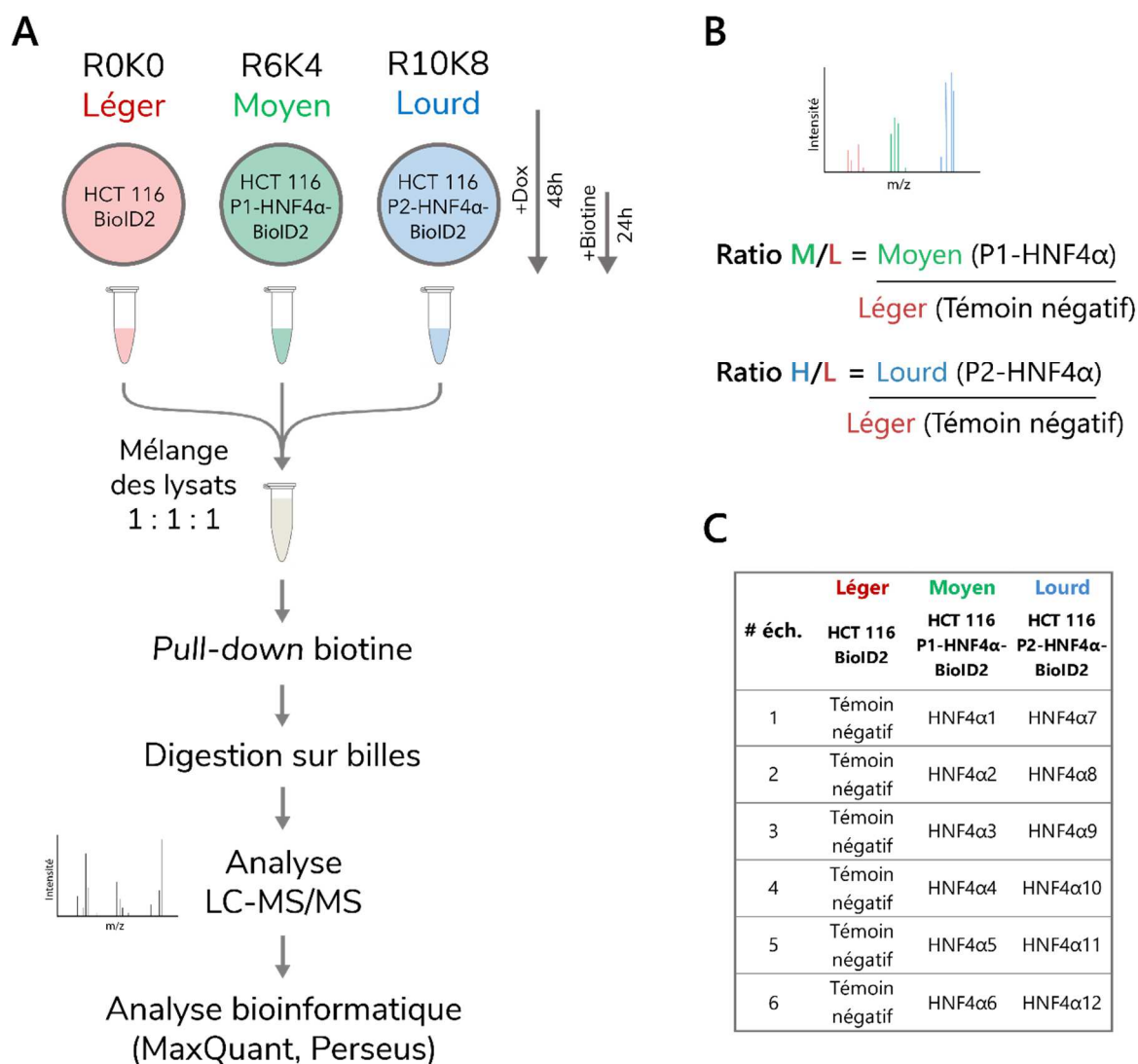


Figure 15 : Stratégie d'identification des partenaires protéiques des isoformes de HNF4α par l'approche BioID couplée à la spectrométrie de masse quantitative. (A) Préparation des échantillons pour l'analyse protéomique. Les lignées HCT 116 HNF4α(1-12)-BioID2-3Xmyc et la lignée contrôle HCT 116 BioID2-3Xmyc-vide ont été cultivées dans du milieu SILAC, induites pendant 48 heures avec la doxycycline et incubées pendant 24 heures avec la biotine. Les extraits totaux provenant de ces cellules ont été combinés à part égale en fonction de leur concentration protéique. Les protéines biotinylées ont été précipitées à l'aide de billes couplées à la streptavidine, puis ont subi une digestion tryptique sur billes. Les peptides ont été analysés par LC-MS/MS, avant de procéder à la quantification des résultats à l'aide du logiciel MaxQuant 1.6.2.2 (n = 3). **(B)** Schéma présentant les ratios d'enrichissement en fonction des conditions expérimentales utilisées pour la culture en milieu SILAC. **(C)** Tableau détaillant les combinaisons des différentes lignées cellulaires pour les milieux légers, moyens et lourds.

comparées, soit la lignée contrôle dans le milieu léger (R0K0), une lignée exprimant une isoforme P1 dans le milieu moyen (R6K4) et une lignée exprimant une isoforme P2 dans le milieu lourd (R10K8). Les extraits de protéines marquées par la biotine provenant de ces trois conditions ont ensuite été mélangés dans un ratio 1 : 1 : 1. Les protéines biotinylées ont subi une digestion par la trypsine, et les peptides résultants ont été analysés par spectrométrie de masse. La quantification des protéines identifiées de façon spécifique pour chaque isoforme a été faite en comparant celles-ci aux protéines identifiées dans la condition contrôle (Figure 15B).

Plus d'un millier de protéines ont été identifiées pour chaque isoforme. Un seuil d'enrichissement minimal de 2 a été utilisé pour l'ensemble des analyses de protéomique, réduisant ainsi le nombre de protéines comparées à des listes d'environ 200 interactants pour chaque isoforme. Les ratios d'enrichissement obtenus pour chaque protéine quantifiée ont été comparés entre les isoformes et leurs triplicatas via une corrélation de Spearman. Ces corrélations ont ensuite été représentées en fonction d'une échelle de couleur afin de les visualiser plus facilement (Figure 16). Cette analyse initiale permet d'une part de valider la reproductibilité des résultats obtenus étant donné la forte corrélation entre les triplicatas (0,67 à 0,87). Également, cette représentation fait ressortir certaines différences entre les protéomes associés à chaque isoforme. Nous retrouvons beaucoup de similarités entre les réseaux d'interaction des isoformes partageant le même domaine A/B. Les listes d'interactants potentiels pour les isoformes $\alpha 4$, $\alpha 5$ et $\alpha 6$ semblent être les plus dissimilaires par rapport aux autres isoformes. De façon globale, il apparaît tout de même qu'une proportion considérable des protéines identifiées sont enrichies de manière similaire pour les 12 isoformes. Suite à cette validation, nous avons choisi d'effectuer le reste des analyses à partir de la moyenne des ratios d'enrichissement obtenus à travers les triplicatas. Étant donné l'expression différentielle des isoformes démontrée dans la section 1.2.1, nous avons décidé de normaliser les ratios d'enrichissement par rapport à la médiane.

Afin de déterminer le type de protéines enrichies par l'approche BioID, des analyses d'enrichissement d'annotations GO ont été réalisées à l'aide du logiciel Panther 13.1. Les annotations de processus biologiques (GO-BP) régulés par ces protéines montrent que la plupart d'entre elles ont des fonctions dans la régulation de la transcription, notamment par la modification des histones et le remodelage de la chromatine (Figure 17A). De façon

logique, les protéines associées à ces processus biologiques enrichis sont annotées avec des fonctions moléculaires (GO-MF) attendues pour des facteurs de transcription et des corégulateurs transcriptionnels (Figure 17B). Il n'y a toutefois pas de différences notables entre le protéome des 12 isoformes qui peuvent être identifiées par ce type d'analyse.

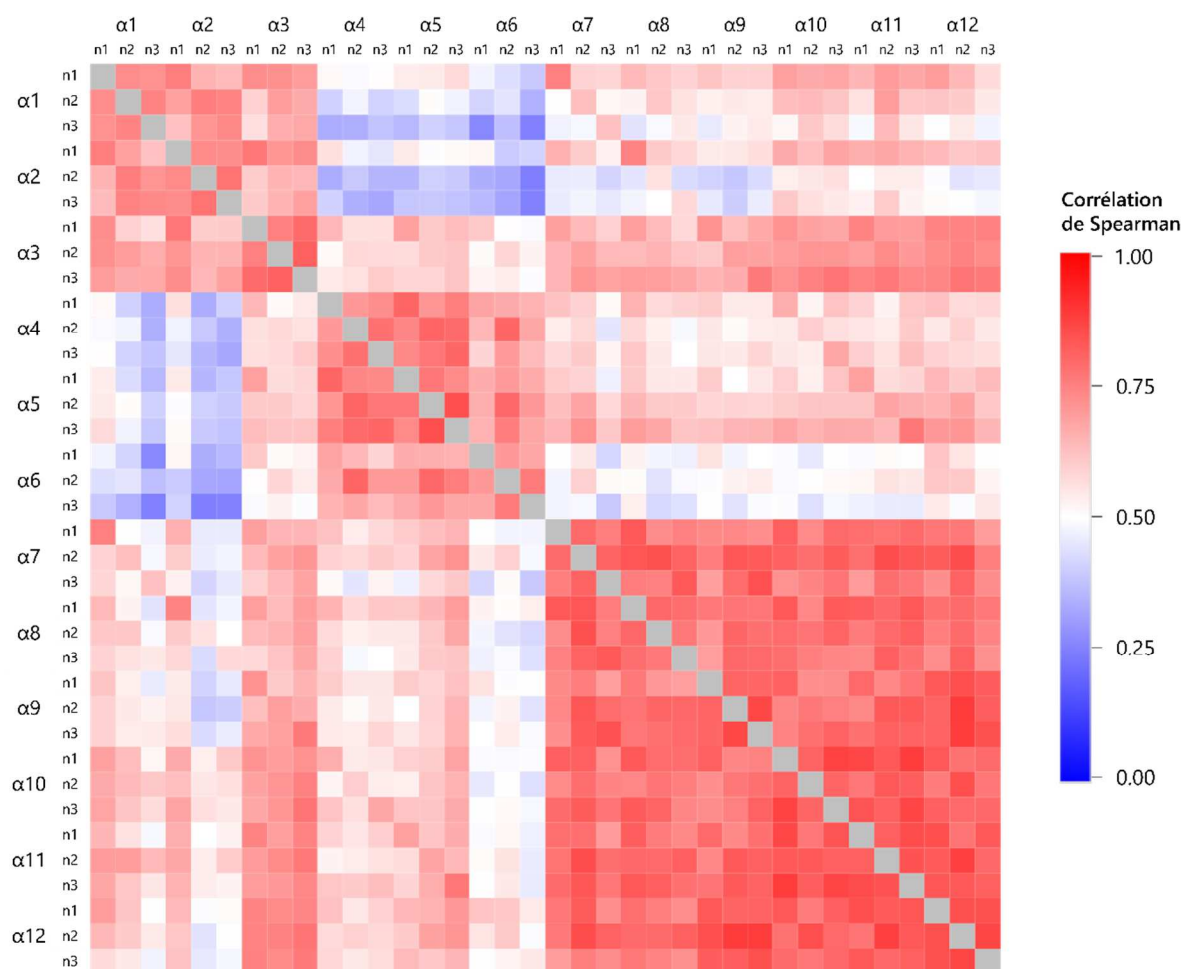


Figure 16 : Le protéome est particulièrement similaire chez les isoformes partageant le même domaine A/B. Visualisation « heatmap » comparant l'association entre les ratios d'enrichissement des protéines identifiées pour chaque isoforme selon une corrélation de Spearman. Les données sont présentées en fonction des triplicatas biologiques effectués pour chaque isoforme. Les principales différences observées regroupent les isoformes qui partagent le même domaine A/B. Cette représentation a été générée à l'aide du logiciel Morpheus. Un seuil minimal de ratio d'enrichissement de 2 a été utilisé.

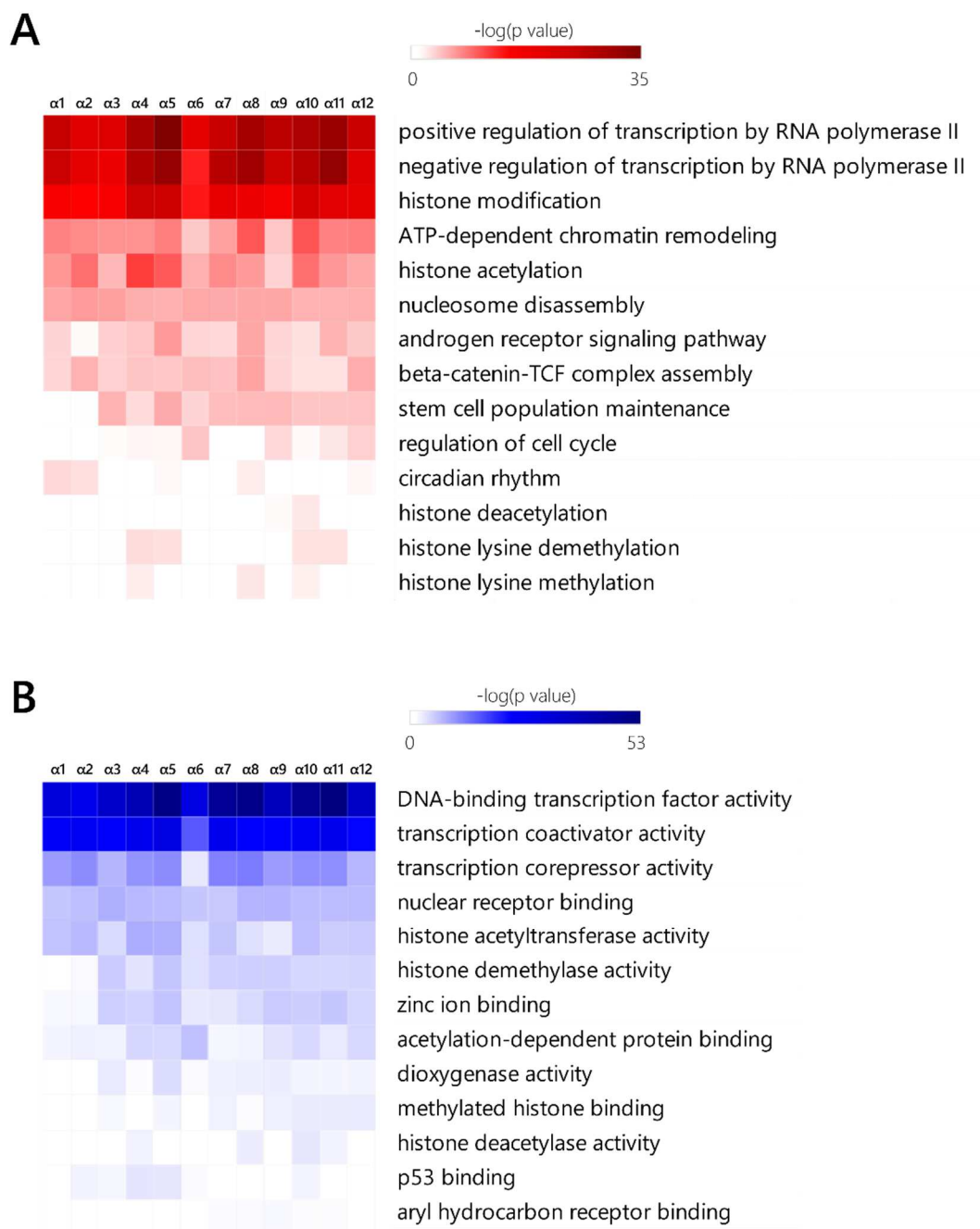


Figure 17 : Les interactants potentiels pour les 12 isoformes sont principalement des corégulateurs transcriptionnels et des facteurs de transcription. Analyse d'enrichissement des annotations **(A)** GO-BP et **(B)** GO-MF pour les protéines identifiées par chacune des isoformes de HNF4α. L'outil Panther 13.1 a été utilisé pour ces analyses. Le logiciel Morpheus a été utilisé afin de visualiser les résultats sous forme de « heatmap ». Un seuil minimal de ratio d'enrichissement de 2 a été utilisé. La valeur négative du logarithme en base 10 du p-value est représentée selon une échelle de couleur pour chaque annotation.

3.2 Identification de partenaires potentiels spécifiques à certaines isoformes de HNF4α

Afin de déterminer si nous pouvions identifier des interactants spécifiques pour certaines isoformes, nous avons choisi de découpler l'effet du domaine A/B et du domaine F. Quatre différents domaines A/B et trois domaines F sont retrouvés parmi toutes les isoformes. En regroupant les protéines identifiées pour toutes les isoformes ayant un de ces domaines en commun, cette analyse a permis de fournir des pistes dans l'identification de protéines pouvant interagir de façon spécifique avec certaines isoformes de HNF4α à travers ces domaines.

Les diagrammes de Venn présentés à la figure 18 illustrent les résultats provenant de cette analyse comparative, qui utilisait un seuil d'enrichissement minimal de 2. De façon globale, nous notons une nouvelle fois que la plupart des protéines identifiées sont enrichies de manière égale entre les 12 isoformes de HNF4α. Parmi elles, on retrouve plusieurs corégulateurs et facteurs de transcription connus pour interagir avec HNF4α, tels CBP (Dell et Hadzopoulou-Cladaras, 1999), NCoA-1 et NCoA-2 (Martínez-Jiménez *et al.*, 2006), NCoR2 (Ruse *et al.*, 2002) et FOSL1 (FRA-1 ; Vuong *et al.*, 2015). Cette analyse fait toutefois ressortir certaines protéines semblant interagir de manière spécifique avec des sous-groupes d'isoformes qui possèdent un domaine A/B ou un domaine F commun. La plupart d'entre elles n'ont jamais été montrées pour interagir spécifiquement avec HNF4α, mais leurs fonctions dans la régulation transcriptionnelle soutiennent cette possibilité. Parmi elles, nous retrouvons des membres de complexes de remodelage de la chromatine dépendants de l'ATP (BRD9 - SWI/SNF ; VPS72 - NuA4 ; SRCAP), des enzymes associées à la modification des histones (SIRT1, SIN3B, KDM2A, KANSL1), ou encore des membres du complexe *Mediator* et du PIC (MED26, MED27, TAF4B).

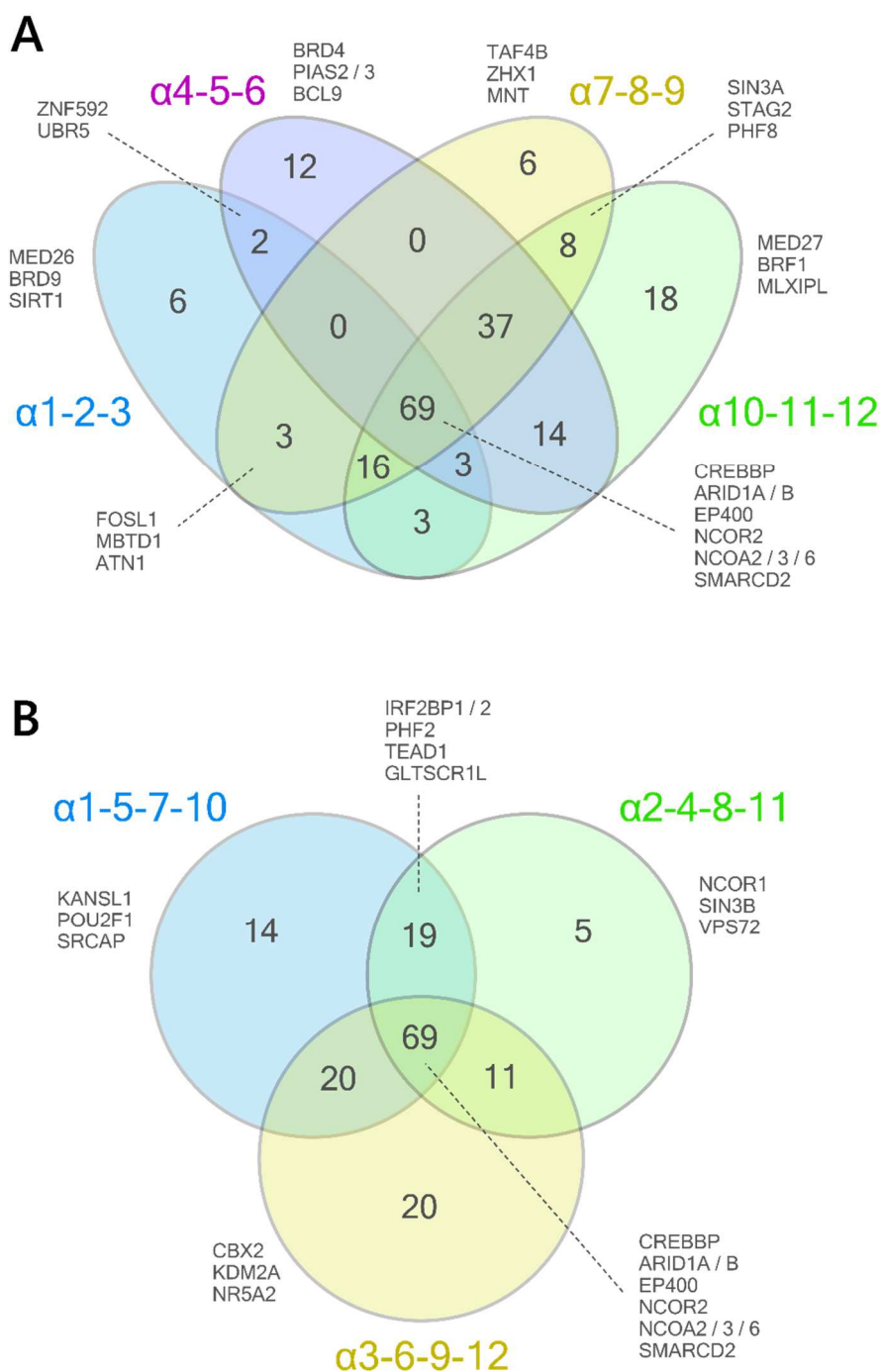


Figure 18 : Plusieurs protéines interagissent spécifiquement avec les isoformes qui possèdent un domaine A/B ou F commun. Diagrammes de Venn comparant les protéines identifiées spécifiquement chez certains sous-groupes d'isoformes de HNF4α. Un seuil minimal de ratio d'enrichissement de 2 a été utilisé. **(A)** Comparaison des interactants spécifiques aux isoformes partageant le même domaine A/B. **(B)** Comparaison des interactants spécifiques aux isoformes partageant le même domaine F.

3.3 Validation des interactions (Co-IP)

L'approche de biotinylation dépendante de la proximité (BioID) couplée à la spectrométrie de masse quantitative nous a permis d'identifier plusieurs interactions potentielles impliquant les différentes isoformes de HNF4α. Cependant, cette approche ne permet pas de discriminer une interaction réelle entre deux protéines par rapport à un événement particulier rapprochant les protéines sans qu'elles n'interagissent entre elles. Plusieurs protéines interagissant potentiellement avec les isoformes ont été sélectionnées afin de valider ces interactions. Le choix des interactions investiguées reposait sur une catégorie de critères objectifs (ratio d'enrichissement par rapport à l'ensemble des protéines, profil d'enrichissement variable selon les sous-groupes d'isoformes) et d'autres critères plus subjectifs (implication potentielle dans la régulation transcriptionnelle par HNF4α, interaction rapportée dans la littérature). La figure 19 montre la sélection initiale de 20 protéines interagissant potentiellement avec HNF4α, ainsi que leur profil d'enrichissement vis-à-vis chacune des 12 isoformes. De ces 20 protéines, 12 d'entre elles ont pu être clonées via RT-PCR et fusionnées à une étiquette FLAG.

Des co-transfections entre ces constructions et les isoformes montrant le plus fort ratio d'enrichissement SILAC (fusionnées à GFP) ont été réalisées afin de confirmer ces interactions. Suite à une immunoprécipitation contre GFP, des essais de co-immunoprécipitation contre FLAG ont permis de valider les interactions impliquant cinq protéines, soit GATAD2B (α11), HNF4γ (α4), IRF2BP2 (α2), MTA1 (α11) et ZNF629 (α11) (figure 20). La condition contrôle, qui n'exprimait que la construction FLAG, confirme la spécificité d'interaction entre HNF4α et ces protéines. Les interactions pour les sept autres protéines ne sont pas pour autant réfutées, puisque ces essais de co-immunoprécipitation ont été réalisés dans des conditions uniformes. Il est donc possible que certaines d'entre elles soient observables suite à une optimisation des conditions pour ces interactions. Néanmoins, cette approche a permis de démontrer de façon non biaisée que l'identification d'interactions par BioID couplée à la quantification SILAC fournit des informations fidèles quant aux réseaux d'interactions de facteurs de transcription comme HNF4α.

	Interactants potentiels	Gènes clonés	Interactants validés
$\alpha 1$ à $\alpha 12$ (++)	ARID1A	ARID1A	ARID1A
	ARID1B	ARID1B	ARID1B
	ARID2	ARID2	ARID2
	GATAD2B	GATAD2B	GATAD2B
	SMARCA4	SMARCA4	SMARCA4
	SMARCC1	SMARCC1	SMARCC1
	SMARCD2	SMARCD2	SMARCD2
	TRIM24	TRIM24	TRIM24
$\alpha 1$ à $\alpha 12$ (+)	BRD8	BRD8	BRD8
	MTA1	MTA1	MTA1
	SIN3A	SIN3A	SIN3A
$\alpha 1$ -2-3	EPC1	EPC1	EPC1
	FOSL1	FOSL1	FOSL1
	ING3	ING3	ING3
	JUN	JUN	JUN
$\alpha 4$ -5-6	HNF4G	HNF4G	HNF4G
	NCOR2	NCOR2	NCOR2
$\alpha 3$ -6-9-12	IRF2BP2	IRF2BP2	IRF2BP2
$\alpha 4$ à $\alpha 12$	PHF21A	PHF21A	PHF21A
	ZNF629	ZNF629	ZNF629

Figure 19 : Listes de protéines sélectionnées et validées pour leur interaction avec les isoformes de HNF4a. 20 interactants potentiels ont été sélectionnés afin de valider leur identification par l'approche BioID couplée à la spectrométrie de masse quantitative. La sélection de ces protéines était basée principalement sur le profil d'interaction préférentielle de celles-ci avec différentes isoformes de HNF4a. De cette liste initiale, 12 gènes ont été clonés et fusionnés à une étiquette FLAG. La co-immunoprécipitation de 5 protéines a permis de confirmer l'interaction de celles-ci avec l'isoforme montrant le plus fort ratio d'enrichissement obtenu par spectrométrie de masse quantitative.

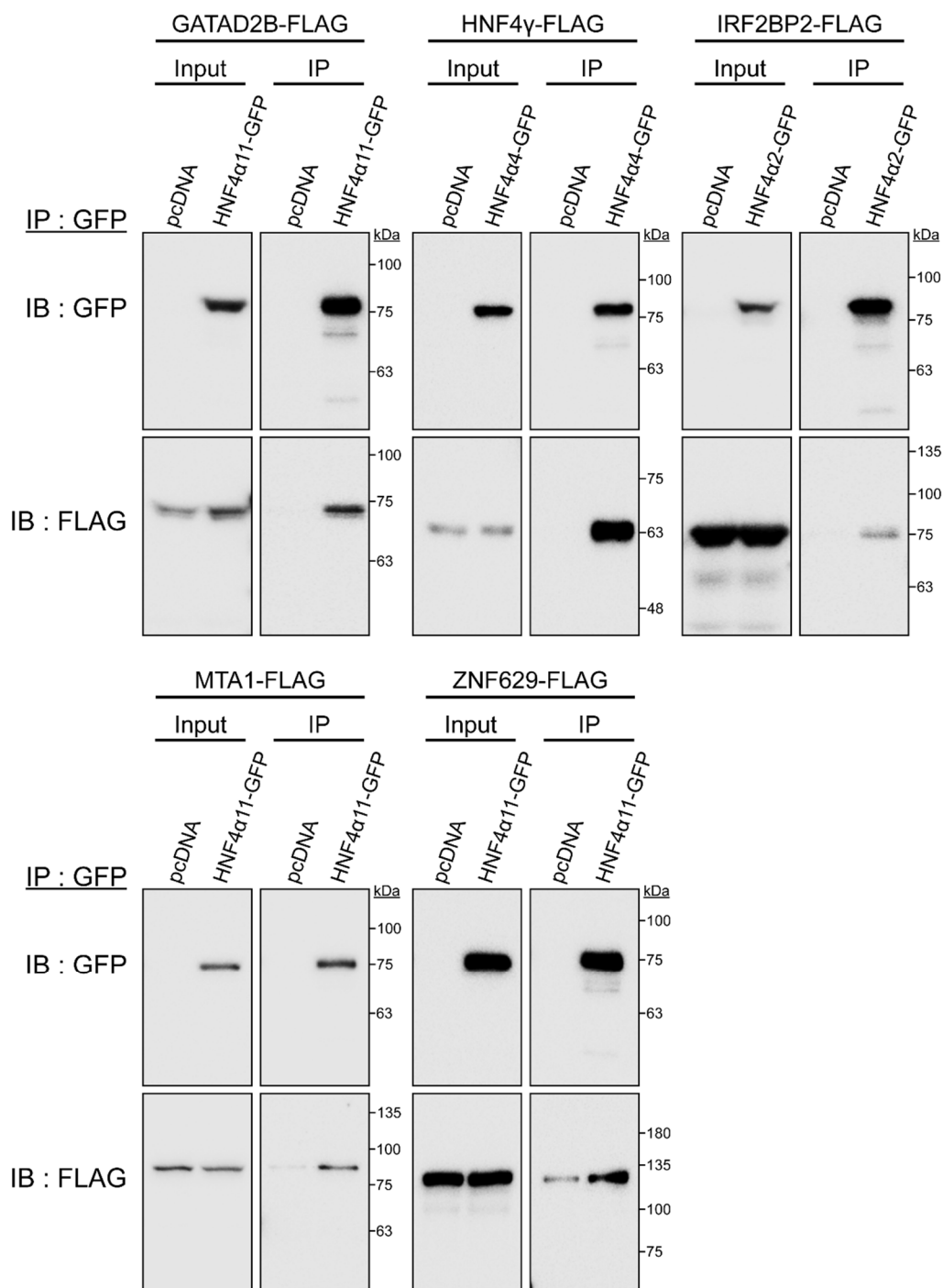


Figure 20 : Validation des interactions entre les isoformes de HNF4 α et cinq protéines identifiées par l'approche BioID. Des cellules 293T ont été co-transfectées pendant 48 heures avec une isoforme de HNF4 α fusionnée à GFP et un partenaire potentiel fusionné à FLAG (GATAD2B, HNF4 γ , IRF-2BP2, MTA1 ou ZNF629). Les complexes protéiques ont été immunoprécipités via GFP à partir d'extraits totaux de ces cellules. La co-immunoprécipitation a été dirigée contre FLAG, permettant de valider l'interaction entre HNF4 α et les cinq protéines indiquées ci-haut (n = 2).

4. Identifier des corégulateurs spécifiques à certaines isoformes pour expliquer les variations dans leurs fonctions transcriptionnelles

4.1 Le corépresseur IRF-2BP2 interagit préférentiellement avec les isoformes de HNF4α possédant la forme longue du domaine F

Les protéines IRF-2BP1 et IRF-2BP2 ont été identifiées initialement comme des corépresseurs du facteur de transcription IRF-2, qui joue un rôle de régulateur négatif dans plusieurs voies de signalisation comme la réponse à l'interféron, le cycle cellulaire et l'apoptose (Childs et Goodbourn, 2003). Ces deux corépresseurs partagent un motif en doigt de zinc C4 et un domaine « RING finger » C3HC4 très conservés. IRF-2BP1 agit aussi en tant que corépresseur pour le FT ATF-2 (*Activating transcription factor 2*) (Kimura, 2008). IRF-2BP2 a été montré plus tard pour être également un corépresseur des FT p53 et NFAT (*Nuclear factor of activated T-cells*) (Koeppel *et al.*, 2009 ; Carneiro *et al.*, 2011).

Dans le contexte de ce projet de recherche, les corépresseurs IRF-2BP1 et IRF-2BP2 ont été identifiés en tant que partenaires potentiels des isoformes de HNF4α. Le profil d'enrichissement de ces corégulateurs en fonction des 12 isoformes était particulier, puisqu'ils avaient été identifiés seulement pour les isoformes possédant la forme longue du domaine F (Figure 21). Les isoformes α3, α6, α9 et α12, qui ont toutes la forme plus courte du domaine F en commun, n'ont pas permis l'identification de ces corépresseurs par l'approche BioID. L'interaction entre IRF-2BP2 et HNF4α2 a précédemment été confirmée par co-IP (Figure 20). L'implication de ces interactions spécifiques à un sous-groupe d'isoformes peut ensuite être intégrée aux analyses transcriptomiques effectuées auparavant. Par souci de simplification, nous nous sommes concentrés sur le profil différentiel observé entre les isoformes α1 et α2 par rapport à l'isoforme α3. Une différence importante au niveau du nombre de gènes modulés négativement par ces isoformes a pu être observée (Figure 13). HNF4α3 est associé à une baisse d'expression pour beaucoup moins de gènes que α1 et α2. La seule différence dans la structure de ces isoformes se situe au niveau de leur domaine F. Ainsi, nous avons avancé l'hypothèse que l'interaction spécifique des corépresseurs IRF-2BP1 et IRF-2BP2 avec les isoformes α1 et α2 pourrait expliquer en partie cette différence au niveau de leur fonction dans la régulation négative

de la transcription. Afin de tester cette hypothèse, nous avons testé de façon préliminaire une approche d'interférence par ARN pour réduire l'expression de la protéine IRF-2BP2 (Figure 22). En affectant l'expression de ce corépresseur, nous serons en mesure de valider si cette modulation a un impact fonctionnel au niveau de la répression transcriptionnelle régulée spécifiquement par les isoformes $\alpha 1$ et $\alpha 2$.

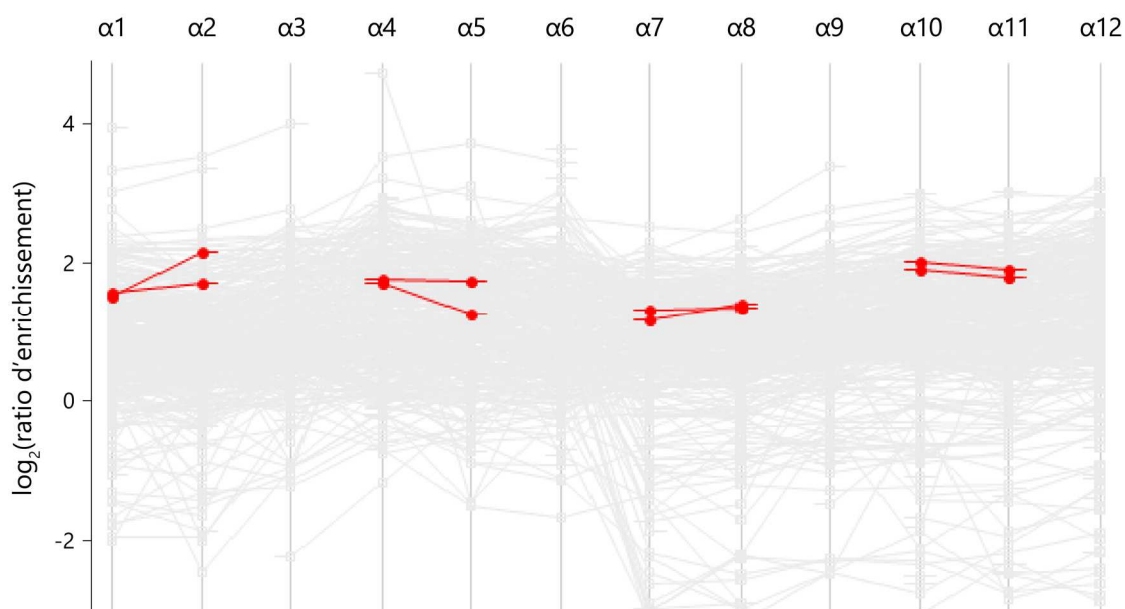


Figure 21 : Les corépresseurs IRF-2BP1 et IRF-2BP2 sont enrichis spécifiquement par les isoformes qui ne possèdent pas la forme courte du domaine F. Profil d'interaction vu en spectrométrie de masse des protéines IRF-2BP1 et IRF-2BP2 avec les 12 isoformes de HNF4 α . Le ratio d'enrichissement obtenu grâce à l'approche BioID dans les lignées HCT 116 HNF4 α (1-12)-BioID2-3Xmyc est présenté sous une forme logarithmique en base 2. L'enrichissement des protéines IRF-2BP1 et IRF-2BP2 est représenté par les points et traits rouges ($n = 3$).

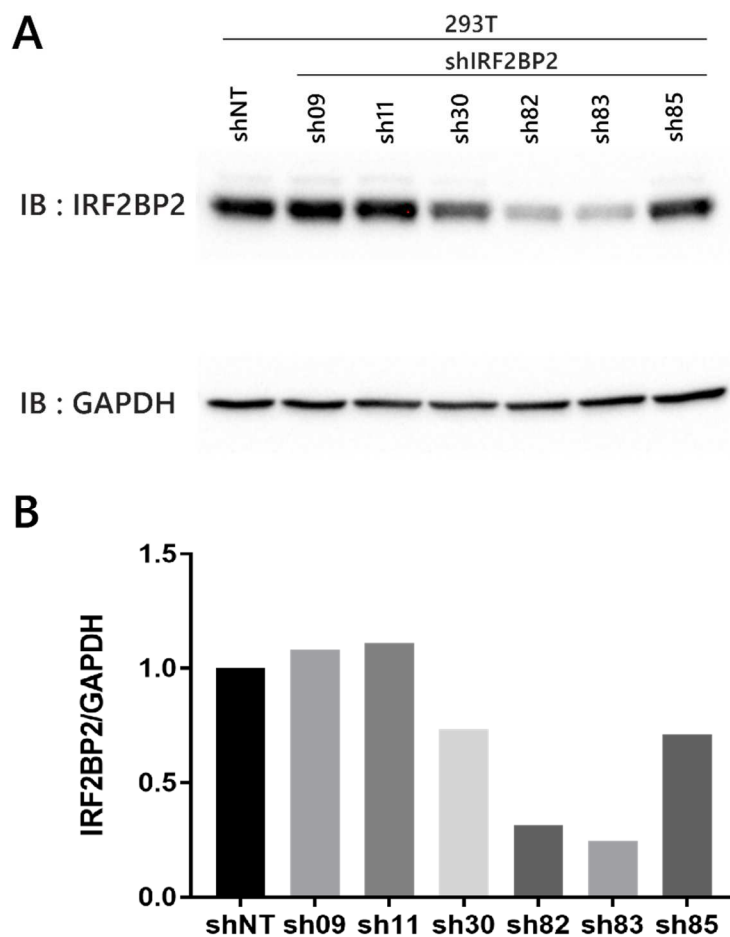


Figure 22 : Validation de l'inhibition du corépresseur IRF-2BP2 via interférence par ARN.

Six shARN dirigés spécifiquement contre IRF-2BP2 ont été testés dans les cellules 293T. Un shARN non spécifique (shNT) a été utilisé comme témoin négatif. **(A)** L'immunobuvardage de la protéine IRF-2BP2 révèle l'efficacité des sh82 et sh83 dans l'inhibition du corépresseur. **(B)** Quantification de la perte d'expression d'IRF-2BP2 par densitométrie relative par rapport à l'expression de la protéine de référence GAPDH, et normalisée par rapport à la condition contrôle (n = 1).

DISCUSSION

1. Le profil d'expression de HNF4 α est hautement variable entre ses 12 isoformes

1.1 Distribution tissulaire des isoformes

HNF4 α est connu pour être exprimé principalement dans le foie, les reins, le pancréas et l'intestin. L'étude la plus complète de la distribution des classes d'isoformes à travers les tissus humains a utilisé des anticorps dirigés contre le domaine A/B spécifique des isoformes P1- et P2-HNF4 α (Tanaka *et al.*, 2006). Les auteurs affirmaient être en mesure de reconnaître les isoformes α 1 à α 6 avec leur anticorps dirigé contre la classe P1. Toutefois, comme détaillé dans l'annexe B, la séquence du domaine A/B des isoformes α 4, α 5 et α 6 utilisée jusqu'à maintenant à travers la littérature n'est pas la bonne. En réalité, le domaine A/B de ces isoformes est complètement différent de celui des isoformes α 1, α 2 et α 3, et ne pouvait donc pas être reconnu dans l'étude de Tanaka et collaborateurs. Les isoformes α 10, α 11 et α 12 possèdent elles aussi un domaine N-terminal différent des autres isoformes P2, et n'étaient donc pas considérées dans cette étude. Comme les anticorps reconnaissaient un domaine A/B commun à trois isoformes, il est également impossible de déterminer l'expression relative de chacune d'entre elles. Ainsi, très peu d'information est présentement disponible quant à la distribution spécifique des isoformes individuelles de HNF4 α .

Nous avons donc voulu initialement déterminer le niveau d'expression spécifique des 12 isoformes à travers différents tissus humains et lignées cancéreuses afin d'éclaircir la situation. Nous faisons face à une limitation majeure, soit le fait que les séquences distinguant les isoformes de façon spécifique sont très courtes et souvent riches en nucléotides G/C. La possibilité de générer des amorces spécifiques permettant une amplification efficace de ces séquences était donc extrêmement limitée. La quantification du niveau d'expression était également impossible compte tenu de la longueur des amplicons qui devaient être produits afin d'être spécifique à une seule isoforme. L'approche de PCR semi-quantitative utilisée malgré ces limitations nous a permis d'identifier de façon

claire les six isoformes canoniques de HNF4 α à travers plusieurs tissus sains et lignées cancéreuses. L'isoforme $\alpha 4$ a également été identifiée, son expression étant restreinte au niveau du foie dans le contexte de ces expériences. Des bandes correspondant à certaines autres isoformes ont pu être observées, mais de façon trop faible relativement au niveau d'amplification non spécifique pour pouvoir conclure quoique ce soit quant à leur expression.

Le fait que nous n'ayons pas réussi à identifier cinq des 12 isoformes dans ces essais pourrait nous mener à remettre en cause la pertinence de poursuivre l'étude fonctionnelle de ces isoformes. Il faut toutefois considérer plusieurs facteurs pouvant expliquer pourquoi ces isoformes n'ont pas pu être toutes identifiées. Premièrement, plusieurs tissus connus pour exprimer HNF4 α n'ont pas été analysés, comme le pancréas et le rein. L'isoforme $\alpha 4$ a initialement été identifiée au niveau du rein, alors que les isoformes $\alpha 10$, $\alpha 11$ et $\alpha 12$ ont été montrées pour être exprimées au niveau des îlots pancréatiques (Drewes *et al.*, 1996 ; Huang *et al.*, 2009). De plus, ces études des isoformes non canoniques utilisaient des amorces reconnaissant simultanément $\alpha 4$, $\alpha 5$ et $\alpha 6$ ainsi que $\alpha 10$, $\alpha 11$ et $\alpha 12$. La détection de ces isoformes n'était donc pas spécifique, mais offrait une meilleure sensibilité d'amplification. Les isoformes de HNF4 α sont également connues pour être impliquées au niveau du développement embryonnaire et de la morphogénèse de plusieurs tissus. L'expression des isoformes canoniques P1 et P2 est modulée de façon importante dans le foie et le pancréas au cours du développement (Torres-Padilla *et al.*, 2001 ; Harries *et al.*, 2008). Il est donc possible que certaines isoformes de HNF4 α aient un rôle spécifique dans ce processus, et qu'elles ne soient plus exprimées chez l'adulte. L'investigation par RT-PCR au niveau d'une plus grande diversité de tissus adultes et fœtaux devrait permettre d'en savoir plus quant à la distribution des 12 isoformes de HNF4 α . Malgré tout, il est également important de considérer l'expression protéique des isoformes, puisqu'elle constitue un déterminant majeur de leur fonction *in vivo*.

1.2 Expression protéique des isoformes

L'analyse de l'expression des isoformes de HNF4 α de manière spécifique est malheureusement limitée à la mesure des niveaux d'expression des transcrits par RT-PCR. Il est en effet impossible d'évaluer de façon spécifique l'expression en protéines des 12 isoformes endogènes, puisque les anticorps reconnaissant HNF4 α ne peuvent pas

reconnaître la combinaison d'un domaine A/B et d'un domaine F spécifique à chaque isoforme. La seule amélioration qui pourrait être proposée quant à l'identification des isoformes endogènes serait la production d'anticorps pouvant reconnaître chacun des domaines A/B et F retrouvés pour HNF4α. Il serait ainsi au moins possible de détecter l'expression de chacun des sous-groupes d'isoformes. Afin de pallier à ce problème lors de ce projet de recherche, nous avons fusionné les isoformes de HNF4α avec les étiquettes protéiques GFP et BioID2-3Xmyc. Bien que cette approche ne nous ait pas permis de conclure quoique ce soit sur l'expression protéique endogène des isoformes, nous avons toutefois pu observer que les isoformes étaient modulées de façon différentielle au niveau post-transcriptionnel. Alors que les niveaux de transcrits étaient similaires entre les différentes isoformes exprimées de façon hétérologue, les niveaux protéiques observés étaient beaucoup plus variables. Nous notons d'abord un profil spécifique d'expression des isoformes en fonction de leur domaine A/B. Certains sous-groupes d'isoformes, comme α4, α5 et α6 sont exprimées à des niveaux protéiques beaucoup plus faibles comparativement aux isoformes α7, α8 et α9, par exemple. Nous constatons également que le domaine F plus court corrèle avec des niveaux d'expression plus faibles pour les isoformes qui contiennent ce domaine (α3, α6, α9 et α12). Curieusement, les extraits nucléaires préparés pour l'essai de retardement sur gel ne présentent pas de telles variations. Nous avons avancé quelques hypothèses pour expliquer ce phénomène, mais aucune d'entre elles n'a pu être validée jusqu'à maintenant. Nous avons montré que la localisation des isoformes est exclusivement nucléaire, et aucun signal pour HNF4α n'était détecté dans les fractions cytoplasmiques testées comme contrôle (données non présentées). Ainsi, nous n'attribuons pas cette variation des niveaux protéiques à la technique d'extraction nucléaire. Il est possible que la surexpression des isoformes dans les cellules 293T compense la dégradation des isoformes plus faiblement exprimées dans les lignées stables. Également, puisque ces extraits nucléaires proviennent d'une lignée cellulaire 293T, contrairement aux lignées stables HCT 116, il n'est pas impossible que les mécanismes responsables de la variation d'expression des isoformes ne soient pas présents dans les cellules 293T. Somme toute, cette observation concerne seulement l'essai de retardement sur gel, les autres objectifs de ce projet ayant été réalisés à l'aide des lignées stables HCT 116.

Puisque les niveaux des transcrits semblent similaires et n'expliquent pas cette variation au niveau des protéines dans les lignées stables, il est possible que la traduction

ou la stabilité des isoformes soient différenciellement affectées. Plusieurs modifications post-traductionnelles sont connues pour réguler la stabilité des protéines. Plus particulièrement pour HNF4 α , il a été montré que les isoformes P1 canoniques étaient régulées négativement par la tyrosine kinase Src. Celle-ci phosphoryle la tyrosine en position 14 qui est retrouvée uniquement dans le domaine A/B des isoformes α 1, α 2 et α 3, ce qui mène à leur déstabilisation et à une perte de localisation nucléaire (Chellappa *et al.*, 2012). De plus, un site de SUMOylation a été identifié au niveau de l'extrémité C-terminale de HNF4 α , et cette modification est associée avec une diminution de la stabilité de la protéine (Zhou *et al.*, 2012). Différentes MPT spécifiques aux domaines A/B et F des isoformes pourraient donc expliquer les variations observées au niveau de leur expression protéique. Afin de valider si ces variations sont dues à des défauts de stabilité protéique, il serait possible d'utiliser un inhibiteur de la traduction comme le cycloheximide pour suivre la demi-vie des isoformes. L'inhibiteur du protéasome MG132 pourrait également être utilisé afin de préciser le mécanisme responsable de la dégradation des isoformes. En lien avec ce dernier point, il serait également intéressant de réanalyser les données de protéomique en fonction des MPT identifiées par spectrométrie de masse. Nous n'avons pas tenu compte de celles-ci initialement dû au temps d'analyse considérablement accru qu'elles nécessitent. Toutefois, plusieurs informations additionnelles quant à des modifications telles que la phosphorylation, l'ubiquitination, l'acétylation ou la méthylation pourraient être obtenues ainsi. Nous pourrions ensuite diriger notre attention sur les MPT spécifiques au niveau des domaines A/B ou F de certaines isoformes, afin de déterminer si ces MPT pourraient affecter leur stabilité. L'étude plus approfondie des variations dans l'expression protéique des différentes isoformes de HNF4 α sera importante pour des considérations relatives à la fonction biologique. Les conclusions tirées dans ce projet quant aux fonctions transcriptionnelles des isoformes ne seront évidemment pertinentes que si ces isoformes sont exprimées à des niveaux endogènes suffisants pour l'exécution de leurs fonctions.

2. Les 12 isoformes de HNF4a modulent de façon distincte le transcriptome des cellules HCT 116

Plusieurs études portant sur HNF4a ont commencé depuis peu de temps à s'intéresser aux fonctions spécifiques de ses classes d'isoformes P1 et P2. Notamment, notre laboratoire a récemment montré des rôles opposés de P1- et P2-HNF4a dans le contexte du cancer colorectal (Babeu *et al.*, 2018). Toutefois, personne ne s'était encore intéressé à comparer de façon individuelle l'ensemble des isoformes composant ces deux classes. Afin d'élucider les fonctions transcriptionnelles spécifiquement régulées par les 12 isoformes de HNF4a, nous avons séquencé le transcriptome des cellules HCT 116 exprimant chacune d'entre elles. L'analyse des résultats de séquençage ayant été débutée tardivement dans ce projet de recherche, nous nous sommes concentrés sur une présentation globale des résultats obtenus. Les analyses préliminaires effectuées jusqu'à maintenant ont cependant fourni des pistes intéressantes pour la suite du projet, puisque des différences majeures ont été observées au niveau des fonctions transcriptionnelles des 12 isoformes. Nous avons montré que les isoformes régulaient l'expression des gènes à des niveaux variables en comparant le nombre de gènes modulés suite à l'expression de chacune d'entre elles. Ces variations peuvent être associées à des sous-groupes d'isoformes partageant le même domaine A/B ou F. HNF4a1 et HNF4a2 sont de loin les plus puissants régulateurs transcriptionnels dans le contexte des cellules HCT 116, puisqu'elles entraînent la modulation de plus d'un millier de gènes dans ce contexte. Il était déjà connu que HNF4a est un régulateur transcriptionnel particulièrement important pour l'expression d'un grand nombre de gènes au sein du même type cellulaire. Notamment, il a été montré que 42 % des gènes occupés par la PolII dans les hépatocytes étaient aussi occupés par HNF4a. Ce taux d'occupation était presque identique au niveau des îlots pancréatiques, soit 43 % (Odom *et al.*, 2004). Les isoformes P2 canoniques $\alpha 7$ - $\alpha 8$ et non canoniques $\alpha 10$ - $\alpha 11$ ont quant à elles un effet plus modeste sur la transcription, puisqu'on retrouve environ de 250 à 500 gènes modulés significativement par ces isoformes. De façon intéressante, nous avons mis en évidence que les isoformes P1 non canoniques ($\alpha 4$, $\alpha 5$ et $\alpha 6$) ont une fonction transcriptionnelle extrêmement limitée. Ces résultats concordent avec l'observation que nous avons faite précédemment concernant la capacité de liaison à l'ADN réduite de ces isoformes. Une

discussion plus approfondie entourant ces isoformes est présentée plus loin dans ce chapitre.

De plus, nous avons noté un impact fonctionnel associé au domaine F de certaines isoformes. Effectivement, il semble que la forme courte du domaine F, qui est présente chez $\alpha 3$, $\alpha 6$, $\alpha 9$ et $\alpha 12$, soit associée à une fonction transcriptionnelle réduite pour ces isoformes par rapport aux isoformes partageant le même domaine A/B. Par exemple, HNF4 $\alpha 3$ entraîne la modulation de seulement 200 gènes, soit une réduction de plus de cinq fois du nombre total de gènes affectés par rapport aux isoformes $\alpha 1$ et $\alpha 2$. Très peu d'études s'étaient intéressées jusqu'à présent à caractériser les isoformes possédant cette forme plus courte du domaine F, puisqu'elles semblent être moins exprimées par rapport aux autres isoformes. L'activité transcriptionnelle de $\alpha 3$ a uniquement été évaluée dans le contexte d'essais luciférase et de retardement sur gel. HNF4 $\alpha 3$ est un transactivateur environ deux fois plus fort pour le promoteur *Hnf1a* murin et le promoteur *APOCIII* humain comparativement à $\alpha 1$ et $\alpha 2$, respectivement (Huang *et al.*, 2008 ; Suaud *et al.*, 1999). Ces observations semblent à première vue contradictoires par rapport à l'activité transcriptionnelle réduite de $\alpha 3$ que nous rapportons dans ce projet. Toutefois, plusieurs facteurs peuvent expliquer ces différences. Ces études exprimaient les constructions des isoformes par surexpression, alors que le système d'expression Flp-In T-REx que nous avons utilisé permettait des niveaux d'expression beaucoup plus similaires à des niveaux endogènes de HNF4 α retrouvés dans les lignées Caco-2/15 et HT-29 (données non présentées). Ainsi, nous avons noté que l'isoforme $\alpha 3$ était exprimée plus faiblement que les isoformes $\alpha 1$ et $\alpha 2$, alors que les études précédentes ne montraient pas de différences à ce niveau. L'expression plus faible de HNF4 α dans notre modèle d'étude pourrait expliquer la réduction de son impact sur le transcriptome. L'analyse du transcriptome représente évidemment une vision globale de l'activité des isoformes, alors que les essais luciférase étaient limités à l'étude de quelques promoteurs seulement. Le contexte cellulaire n'était également pas le même dans ces différentes études. Pour toutes ces raisons, nous pouvons affirmer que notre étude transcriptomique est la première à démontrer que l'isoforme $\alpha 3$ a une activité transcriptionnelle réduite comparativement aux isoformes $\alpha 1$ et $\alpha 2$.

Les listes de gènes identifiés pour être modulés significativement dans ce projet peuvent être comparées aux résultats obtenus par Vuong et collaborateurs, qui ont analysé

le transcriptome des cellules HCT 116 après y avoir exprimé de façon hétérologue les isoformes $\alpha 2$ et $\alpha 8$ (Vuong *et al.*, 2015). La technologie de séquençage utilisée par cette équipe leur a permis d'identifier un nombre nettement réduit de gènes modulés par rapport à ce que nous avons obtenu. Selon un seuil minimal de modulation de 2, 67 et 85 gènes étaient respectivement modulés par $\alpha 2$ et $\alpha 8$ (contre 1108 et 238 pour ce projet). Néanmoins, la majorité des gènes les plus modulés étaient associés aux mêmes processus biologiques que nous avons pu observer, comme la réponse aux xénobiotiques, l'adhésion cellulaire et la réponse aux stimuli extracellulaires.

Des analyses plus approfondies devront être effectuées afin de déterminer si certaines voies de signalisation cellulaire ou processus biologiques sont modulés spécifiquement par certaines isoformes. Nous pourrions également intégrer à cette étude des expériences de ChIP-seq afin d'identifier les gènes liés par les différentes isoformes. En intégrant ces expériences aux résultats de séquençage de l'ARN obtenus durant ce projet, nous serions ainsi en mesure de déterminer la composition des réseaux de gènes cibles pour chacune des isoformes. De façon globale, l'analyse préliminaire des résultats de transcriptomique nous permet de conclure que des différences majeures sont présentes entre les 12 isoformes de HNF4 α , et sont associées aux domaines A/B et F variables de ces isoformes. L'étude des interactions spécifiques entourant ces différents domaines, détaillée plus loin dans ce chapitre, visait donc à identifier des mécanismes de régulation transcriptionnelle pouvant expliquer ces différences fonctionnelles.

3. La spectrométrie de masse quantitative révèle des réseaux d'interactions spécifiques aux isoformes de HNF4 α

3.1 Les fonctions différentielles des isoformes sont finement régulées via leur interaction avec un sous-ensemble de régulateurs transcriptionnels

Les 12 isoformes de HNF4 α varient seulement au niveau des extrémités N- et C-terminales de leur séquence. Nous avons présumé que ces différences structurelles permettraient aux isoformes d'interagir spécifiquement avec certains régulateurs transcriptionnels, ce qui expliquerait les variations fonctionnelles observées précédemment

au niveau transcriptomique. Afin de vérifier cette hypothèse, nous avons utilisé une approche BioID couplée à la spectrométrie de masse quantitative, qui nous a permis d'établir les réseaux d'interactions protéiques des 12 isoformes. La normalisation par rapport à la médiane des ratios d'enrichissement a dû être effectuée afin de tenir compte des niveaux d'expression différentiels observés pour les isoformes. La comparaison globale des réseaux d'interaction obtenus montre que la majorité du protéome est similaire entre chaque isoforme. Plusieurs profils différentiels peuvent cependant être observés. De manière analogue à ce que nous avons noté précédemment dans les analyses de transcriptomique, le domaine A/B semble coordonner la majorité des interactions spécifiques à certaines isoformes. En effet, nous observons les corrélations les plus fortes entre les réseaux des isoformes qui partagent ce domaine.

Nous avons montré que la vaste majorité des interactants potentiels des isoformes de HNF4 α sont associés à des processus biologiques de régulation transcriptionnelle et de modification de la structure de la chromatine. Nous pouvons conclure que l'utilisation d'un témoin négatif exprimant seulement l'étiquette BioID2 permet de réduire l'identification non spécifique de protéines souvent associée à cette approche lorsque ce type de contrôle n'est pas utilisé. L'approche BioID utilisée selon ces paramètres nous a ainsi permis d'identifier de façon spécifique et sensible plusieurs protéines ayant un fort potentiel d'interaction avec HNF4 α .

3.2 L'approche BioID permet l'identification de plusieurs protéines ayant une réelle interaction avec HNF4 α

Des listes d'environ 200 partenaires potentiels enrichis significativement ont été obtenues pour chaque isoforme de HNF4 α . Parmi ces protéines, certains interactants connus de HNF4 α ont pu être identifiés, incluant les coactivateurs CBP (Dell et Hadzopoulou-Cladaras, 1999), NCoA-1, NCoA-2 et NCoA-3 (Lu, 2016) et le corépresseur NCoR2 (Ruse *et al.*, 2002). Par contre, plusieurs autres corégulateurs bien caractérisés de HNF4 α comme PGC-1 α , p300 et PC4 n'ont pas été détectés lors de notre étude. Il est possible que dans le contexte des cellules HCT 116 que nous avons sélectionnées, ces protéines soient exprimées à des niveaux trop faibles pour pouvoir être identifiées. L'approche BioID que nous avons utilisée était également différente, puisque la plupart des

études ayant caractérisé les partenaires d'interaction de HNF4α réalisaient ces essais en immunoprécipitant directement les protéines étudiées suite à leur surexpression. Ainsi, il est tout à fait plausible de ne pas avoir pu identifier tous les partenaires connus de HNF4α en utilisant une seule approche dans un contexte cellulaire précis.

Puisque l'approche BioID dépend de la proximité d'une protéine pour son identification, nous ne pouvions pas conclure si toutes les protéines observées en spectrométrie de masse étaient réellement des partenaires de HNF4α ou s'il s'agissait simplement de protéines situées dans son environnement rapproché. Afin de valider le choix de cette approche, nous avons sélectionné une courte liste de protéines ayant des profils d'interaction différentiels en fonction des isoformes de HNF4α. Parmi les 12 partenaires potentiels testés par co-immunoprécipitation, nous avons confirmé que cinq d'entre eux interagissaient avec HNF4α. La co-IP de ces protéines a été effectuée par une approche GFP-Trap. Le fait que nous ayons observé ces interactions par deux approches différentes représente un niveau de validation supplémentaire. Nous avons choisi d'effectuer les essais de co-IP dans des conditions uniformes pour les 12 protéines testées. Toutefois, plusieurs interactions protéine-protéine sont très sensibles à différents paramètres comme le pH et la force ionique du tampon utilisé pour l'immunoprécipitation. Il serait donc probablement possible d'optimiser ces paramètres de façon spécifique afin de valider l'interaction pour une partie des autres protéines testées. Néanmoins, le but de cette validation était de démontrer que l'approche BioID nous permettait d'identifier des interactions réelles entre divers corégulateurs et facteurs de transcription et HNF4α, ce qui est effectivement le cas.

4. Le corépresseur IRF-2BP2 pourrait réguler spécifiquement les isoformes ne possédant pas la forme courte du domaine F

Après avoir analysé le transcriptome et le protéome spécifiques à chacune des 12 isoformes de HNF4α, nous avons ensuite tenté d'intégrer ces différents ensembles de données dans le but d'identifier des profils similaires pour certaines isoformes. Il est important de mentionner que ces analyses sont encore très préliminaires, et pour cette raison nous n'avons présenté qu'un seul exemple de protéine que nous souhaiterions étudier pour la suite de ce projet de recherche. Évidemment, plusieurs autres régulateurs potentiels

de la fonction transcriptionnelle des isoformes de HNF4 α ont été identifiés et seront les sujets d'analyses plus approfondies.

Nous avons identifié le corépresseur IRF-2BP2 comme un nouveau corégulateur de HNF4 α . Un autre corépresseur hautement similaire à celui-ci, IRF-2BP1, a également été identifié de la même manière, mais nous référerons seulement à IRF-2BP2 par souci de concision lors de cette discussion. Cette protéine a initialement été montrée pour être un corépresseur du facteur de transcription IRF-2, et plus récemment pour d'autres FT comme p53 et NFAT (Childs et Goodbourn, 2003 ; Koeppel *et al.*, 2009 ; Carneiro *et al.*, 2011). Nous avons identifié IRF-2BP2 comme un partenaire potentiel d'un sous-groupe d'isoformes de HNF4 α . L'approche BioID a montré que les isoformes qui possèdent la forme courte du domaine répresseur F n'interagissaient aucunement avec ce corégulateur. Nous avons initialement validé par co-IP l'interaction d'IRF-2BP2 avec l'isoforme α 2, soit celle montrant le plus fort ratio d'enrichissement pour cette protéine. Par la suite, nous avons tenté de reproduire le profil d'interaction différentiel détecté pour les 12 isoformes selon cette approche. L'essai n'a toutefois pas permis d'observer les mêmes différences, puisque la protéine IRF-2BP2 interagissait de façon similaire avec les 12 isoformes dans ce contexte (données non présentées). En rétrospective, l'approche de co-IP que nous avons choisie n'était pas la meilleure méthode pour identifier les mêmes interactions différentielles mesurées par spectrométrie de masse. Un premier problème était le fait que nous avons surexprimé à la fois l'isoforme de HNF4 α ainsi que IRF-2BP2. Par contraste avec l'approche BioID, IRF-2BP2 était identifiée à partir de ses niveaux endogènes suite à l'expression modérée de l'isoforme de HNF4 α . Ainsi, l'augmentation importante de l'expression des deux protéines étudiées limitait la possibilité de reproduire la détection des variations observées à des niveaux endogènes. Également, les conditions de *pull-down* utilisées n'étaient pas les mêmes, puisque l'approche GFP-Trap utilisée pour les essais de co-IP ne pouvait pas être réalisée dans des conditions aussi stringentes que pour l'approche BioID. Comme mentionné précédemment, ces conditions variables peuvent avoir un effet important sur l'efficacité de la précipitation des protéines. Afin de contourner ces limitations, nous pourrions refaire les expériences de co-IP en nous servant des lignées HCT 116 HNF4 α (1-12)-GFP pour retourner à des niveaux d'expression similaires à ceux utilisés lors de l'approche BioID. Nous pourrions utiliser un anticorps dirigé contre IRF-2BP2 pour sa détection en co-IP, ce qui nous permettrait d'éviter d'avoir à surexprimer la protéine en fusion avec une étiquette FLAG.

En tenant compte du profil différentiel d'interaction de IRF-2BP2 avec les isoformes de HNF4a, nous avons identifié des variations au niveau du transcriptome pouvant être expliquées par ces interactions préférentielles. Nous avons montré que la proportion de gènes modulés à la baisse était similaire entre les isoformes $\alpha 1$ et $\alpha 2$ (34 % et 37 % respectivement). Cette proportion chutait à 21 % au niveau de l'isoforme $\alpha 3$, avec au total environ 10 fois moins de gènes modulés négativement. Comme nous avons démontré par l'approche BioID que le corépresseur IRF-2BP2 interagissait spécifiquement avec les isoformes $\alpha 1$ et $\alpha 2$ par rapport à $\alpha 3$, nous avons émis l'hypothèse que IRF-2BP2 serait partiellement responsable de cette variation dans la régulation transcriptionnelle négative de ces isoformes. Afin de tester cette hypothèse, nous souhaiterions utiliser une stratégie d'interférence par ARN dirigée contre IRF-2BP2. S'il s'avère être un corépresseur spécifique des isoformes $\alpha 1$ et $\alpha 2$, la perte de son expression devrait être associée à un nombre réduit de gènes modulés négativement par ces isoformes. L'isoforme $\alpha 3$ servirait dans ce contexte de condition contrôle, et ne devrait pas être affectée par la perte d'expression d'IRF-2BP2. Différentes stratégies pourront être utilisées afin de déterminer les changements causés par l'utilisation d'un shARN contre IRF-2BP2. La méthode la plus rapide serait de séquencer l'ARN de ces échantillons et de comparer leur transcriptome aux données déjà obtenues. De façon préliminaire, nous pourrions aussi tenter d'identifier des gènes connus pour être régulés négativement par IRF-2BP2 parmi les gènes réprimés suite à l'expression des isoformes $\alpha 1$ et $\alpha 2$. Certains problèmes relatifs à l'essai d'interférence par ARN peuvent déjà être anticipés. IRF-2BP2 est exprimée en deux isoformes qui semblent être redondantes au niveau fonctionnel (Childs et Goodbourn, 2003). Nous avons cependant confirmé que les shARN ciblant IRF-2BP2 reconnaissent ses deux isoformes par immunobuvardage. Toutefois, il est possible que la perte d'expression d'IRF-2BP2 soit compensée par le corépresseur IRF-2BP1, qui est lui aussi présent dans les cellules HCT 116. La génération de shARN ciblant les domaines à doigts de zinc ou RING hautement conservés entre ces deux corépresseurs pourrait permettre de contrer cet effet de redondance en inhibant leur expression de façon simultanée.

L'hypothèse d'une répression spécifique par IRF-2BP2 en fonction du domaine F est toutefois incohérente à première vue lorsqu'on considère également les autres sous-groupes d'isoformes. Par exemple, l'expression des isoformes $\alpha 7$ et $\alpha 8$ est associée à une proportion de gènes réprimés de 50 % et 41 %, respectivement. Nous observons que

contrairement à l'isoforme $\alpha 3$, $\alpha 9$ qui possède le même domaine F est plutôt associée à un taux plus important de gènes réprimés (55 %). Cela nous indique qu'une analyse plus en profondeur des gènes régulés par les différentes isoformes est nécessaire afin de savoir si certains d'entre eux sont spécifiquement modulés en fonction des domaines de HNF4 α . D'autre part, une limitation importante de nos analyses actuelles est le fait que nous avons uniquement considéré la structure des isoformes de façon linéaire. En d'autres mots, nous avons choisi de considérer le domaine A/B et le domaine F comme des entités indépendantes, puisqu'elles sont situées aux extrémités N- et C-terminales de HNF4 α . Ainsi, nous avons analysé l'effet du domaine F de façon dissociée du domaine A/B en regroupant les isoformes partageant ce domaine, et vice-versa. Cependant, la structure tridimensionnelle de ces domaines variables n'a jamais été déterminée pour HNF4 α , comme pour la majorité des RN, étant donné leur repliement en une structure tertiaire mal définie. Également, il a déjà été montré que les régions activatrices AF-1 et AF-2 de plusieurs RN pouvaient coopérer afin d'activer la transcription en liant le même coactivateur de manière synergique. C'est le cas notamment pour l'interaction de NCoA-2 avec ER α (*Estrogen receptor alpha*) et RAR α (Benecke *et al.*, 2000 ; Bommer *et al.*, 2002). Dans le cas de HNF4 α , il est connu que l'interaction du coactivateur CBP dépend à la fois de la présence des régions AF-1 et AF-2 afin d'atteindre un niveau de transactivation maximal (Dell et Hadzopoulou-Cladaras, 1999). Dans le même ordre d'idée, il est possible que l'interaction du corépresseur IRF-2BP2 ne soit pas suffisante pour moduler négativement l'activité de HNF4 α et qu'elle nécessite le recrutement d'autres corépresseurs via le domaine A/B. En somme, un niveau de complexité important se rajoute lorsque nous tentons d'intégrer à la fois l'effet du domaine A/B et du domaine F spécifiques à chacune des 12 isoformes. Avant tout, la première étape à réaliser dans le futur consistera en une analyse plus poussée des résultats de transcriptomique, qui pourront ainsi nous permettre d'identifier des gènes modulés spécifiquement par certaines isoformes.

5. Plusieurs nouveaux interactants potentiels de HNF4α ont été identifiés de façon indépendante des isoformes exprimées

Bien que dans le cadre de ce projet, nous nous intéressions à définir de nouvelles interactions spécifiques pour certaines isoformes de HNF4α, nous avons également identifié plusieurs protéines enrichies par les 12 isoformes de façon égale. En réalité, la majorité des interactants potentiels identifiés avaient un profil d'enrichissement constant. Cette observation est cohérente avec le fait qu'une grande partie de la séquence de HNF4α est commune aux 12 isoformes, incluant la région activatrice AF-2. Parmi les protéines les plus enrichies et qui ne montraient pas de spécificité vis-à-vis certaines isoformes, nous avons identifié une quantité remarquable de protéines faisant partie de larges complexes de remodelage de la chromatine dépendants de l'ATP. Les complexes de ce type qui étaient les plus surreprésentés étaient BAF (*BRG1/BRM-associated factor*) et PBAF (*Polybromo-associated BAF*) de la famille SWI/SNF, NuRD et NuA4 (*Nucleosome acetyltransferase of H4*).

Le complexe NuRD est un complexe de remodelage des nucléosomes possédant également une activité HDAC, et est généralement associé à des fonctions de répression transcriptionnelle (Basta et Rauchman, 2015). 10 protéines composant NuRD ont été enrichies par les différentes isoformes de HNF4α, sur un total de 12 sous-unités possibles. Le complexe NuA4 possède pour sa part une activité HAT en plus de son activité de remodelage de la chromatine (Jacquet *et al.*, 2016). Des 16 sous-unités formant ce large complexe chez l'humain, nous avons identifié la moitié d'entre elles comme des interactants potentiels de HNF4α. L'identification des membres des complexes SWI/SNF BAF et PBAF est probablement la plus prometteuse compte tenu du grand nombre de sous-unités enrichies à des niveaux très élevés pour l'ensemble des isoformes (Tableau 27, annexe G). Les complexes BAF et PBAF partagent la majorité de ces sous-unités et sont formés minimalement de 14 protéines. Nous avons identifié 10 sous-unités BAF et 11 sous-unités PBAF, pour la plupart enrichies fortement par les 12 isoformes de HNF4α. Même si ces exemples de complexes ne présentent pas de profil d'interaction spécifique à certaines isoformes, l'identification de ces protéines par les 12 isoformes peut être considérée comme une observation de ces interactions ayant été répétée 12 fois pour HNF4α, de manière globale. En tenant compte des triplicatas biologiques que nous avons effectués, cela

représente une identification qui a été réalisée un total de 36 fois pour la majorité des protéines composant ces complexes. Ainsi, malgré le fait que nous n'ayons pas encore validé ces interactions potentielles, nous pouvons avancer avec confiance que ces observations ne sont pas dues à un simple hasard.

Plusieurs autres interactants potentiels pour HNF4 α , comme des membres des complexes corépresseurs SIN3-HDAC et CDYL (*Chromodomain Y like*), ont été identifiés et représentent d'autres types de corégulateurs intéressants à étudier dans le cadre de projets futurs sur la fonction transcriptionnelle de HNF4 α . Aucune interaction impliquant ces différentes protéines n'a été démontrée dans la littérature pour HNF4 α . La validation de ces interactions consisterait donc une avancée importante dans l'étude des mécanismes régulant l'activité transcriptionnelle de HNF4 α .

6. La fonction des isoformes α 4-5-6 est indépendante de leur liaison au site consensus DR1 de HNF4 α

Comme les 12 isoformes de HNF4 α partagent le même domaine de liaison à l'ADN, nous avons émis l'hypothèse que leur capacité à reconnaître les éléments de réponse de HNF4 α était la même dans un contexte indépendant de la chromatine. De façon étonnante, nous avons démontré que les 12 isoformes pouvaient lier le site DR1 consensus de HNF4 α , à l'exception des isoformes α 4, α 5 et α 6. Cette observation n'était pas liée au niveau d'expression protéique, qui était similaire d'une isoforme à l'autre dans les extraits nucléaires utilisés. Les isoformes α 4, α 5 et α 6 possèdent un domaine A/B commun qui n'est présent chez aucune autre isoforme. Nous avons observé que la capacité de liaison diminuée de ces isoformes était associée à une abolition quasi-totale de leur activité transcriptionnelle. Une analyse préliminaire du faible nombre de gènes modulés par ces isoformes ne montre aucun gène cible connu de HNF4 α parmi ceux-ci.

Ces évidences nous ont poussés à comprendre quelle autre fonction pourraient accomplir des variants d'épissage d'un facteur de transcription incapables de réguler directement l'expression de ses gènes cibles. Certains exemples d'isoformes présentant des caractéristiques similaires pour d'autres récepteurs nucléaires ont été identifiés à travers la

littérature. L'isoforme GR β est un variant « dominant négatif » du récepteur aux glucocorticoïdes, puisqu'il antagonise l'activité de l'isoforme active GR α en se liant aux mêmes éléments de réponse sans être capable d'activer la transcription (Oakley *et al.*, 1996). L'épissage alternatif de PR au niveau de son domaine de liaison à l'ADN génère l'isoforme PR-C. La capacité de liaison à l'ADN de cette isoforme est supprimée due à une délétion de son DBD, mais elle régule tout de même la transcription en inhibant directement les isoformes PR-A et PR-B (Wei *et al.*, 1996). Un dernier exemple particulièrement intéressant implique les isoformes du récepteur nucléaire orphelin ROR α (*RAR-related orphan receptor alpha*) (NR1F1). Ses isoformes ROR α 1 et ROR α 2 reconnaissent le même type d'éléments de réponse (RORE), mais selon une spécificité de liaison différente. ROR α 1 active fortement la transcription d'une multitude de gènes, alors que ROR α 2 se lie à une faible diversité de RORE et est un plus faible activateur transcriptionnel. La différence structurelle entre ces deux isoformes se situe dans leur domaine N-terminal, soit la situation exacte entourant les isoformes α 4-6 de HNF4 α . Étonnamment, les capacités de liaison et de transactivation associées aux isoformes ROR α 1 et ROR α 2 sont transférées au récepteur hétérologue TR β lorsque le domaine N-terminal de TR β est substitué par ceux des isoformes de ROR α (Giguère *et al.*, 1994). Cela représente une preuve que le domaine A/B d'un récepteur nucléaire peut réguler sa liaison à l'ADN et ainsi son activité transcriptionnelle. Il serait intéressant d'effectuer le même type d'expérience en transférant le domaine A/B de HNF4 α 4-5-6 au niveau d'un autre récepteur.

En faisant l'intégration de ces différents exemples, nous avons proposé différents mécanismes par lesquels les isoformes HNF4 α 4-6 pourraient moduler l'expression des gènes. Comme la plupart des partenaires d'interaction potentiels de ces isoformes sont des corégulateurs transcriptionnels retrouvés également pour les autres isoformes, il est possible que α 4-6 exercent une fonction de régulation transcriptionnelle en séquestrant ces corégulateurs pour ainsi empêcher leur recrutement à la chromatine. Ces isoformes pourraient aussi réguler la transcription de façon HNF4 α -indépendante via leurs interactions avec d'autres FT. De façon intéressante, nous avons identifié le récepteur nucléaire HNF4 γ (*Hepatocyte nuclear factor 4 gamma*) (NR2A2) en tant que partenaire d'interaction de HNF4 α , avec qui il partage une forte homologie. Nous avons montré par l'approche BioID un profil d'interaction préférentielle de ce RN pour les isoformes α 4-6, et cette interaction a été validée pour HNF4 α 4. HNF4 γ a par ailleurs été identifié récemment comme un FT

capable de transactiver les mêmes gènes que HNF4 α , et même plus fortement dans le cas de son isoforme HNF4 γ 2 (Sasaki *et al.*, 2018). Une possibilité intéressante serait donc que les isoformes HNF4 α 4-6 agissent en tant que corégulateurs pour ce FT. Le récepteur nucléaire SHP, qui ne peut pas lier l'ADN, est un exemple de RN connu pour agir en tant que corégulateur de la transcription (Lee *et al.*, 2000). Une dernière hypothèse que nous avançons en lien avec la fonction des isoformes α 4-6 serait qu'elles puissent se dimériser avec d'autres isoformes de HNF4 α . À travers la littérature, HNF4 α est reconnu comme un homodimère exclusif, mais cette terminologie réfère à la formation de dimères entre différents RN (Jiang *et al.*, 1995). Le concept « d'hétérodimères » formés de deux isoformes distinctes a été très peu étudié chez les récepteurs nucléaires. Il a toutefois été proposé suite à des essais de retard sur gel que les isoformes α 1 et α 7 pouvaient former des dimères l'une avec l'autre (Torres-Padilla *et al.*, 2001).

Le contexte de ce projet de recherche ne permettait pas de vérifier les propriétés de dimérisation des isoformes, puisque chacune d'entre elles était exprimée de façon individuelle dans les cellules HCT 116. Il serait intéressant d'approfondir l'étude de « l'hétérodimérisation » entre différentes isoformes de HNF4 α . L'approche utilisée à cet effet dans l'étude de Torres-Padilla et collaborateurs fournissait des preuves peu convaincantes, compte tenu du fait que la détection de la dimérisation était basée uniquement sur la petite différence de taille entre les deux isoformes. L'utilisation de constructions d'isoformes fusionnées à deux étiquettes protéiques de tailles distinctes (GFP et FLAG, par exemple) permettrait d'améliorer ces essais de retard sur gel, en augmentant la différence de poids moléculaire entre les isoformes et en permettant la reconnaissance spécifique de chacune d'entre elles à l'aide d'un anticorps. Une approche complémentaire serait d'exprimer simultanément dans les cellules HCT 116 une isoforme possédant une faible activité transcriptionnelle, comme α 4, ainsi qu'une isoforme activant fortement la transcription comme α 1. S'il y a dimérisation, l'ajout d'une quantité progressivement plus élevée de l'isoforme α 4 devrait par conséquent réduire l'activation transcriptionnelle par α 1. Cela représenterait une démonstration indirecte de l'impact fonctionnel d'une dimérisation potentielle entre deux isoformes de HNF4 α .

CONCLUSION GÉNÉRALE

HNF4 α est un facteur de transcription important pour le développement et la maintenance de la fonction de plusieurs organes. Récemment, des fonctions opposées entre ses deux classes d'isoformes P1 et P2 ont été identifiées dans différents tissus et pathologies tel le cancer colorectal. Les fonctions individuelles des isoformes composant ces classes ont pour leur part été caractérisées de manière très superficielle jusqu'à maintenant. Ce projet de recherche représente la première étude comparative de l'ensemble des isoformes de HNF4 α . Les résultats obtenus nous ont permis de dresser un portrait global des fonctions transcriptionnelles associées à chacune des 12 isoformes de HNF4 α .

Nous avons démontré des différences majeures dans l'activité transcriptionnelle des isoformes, qui corrélaient avec des variations plus subtiles au niveau de leurs réseaux d'interactions protéiques. Nos résultats suggèrent que le domaine A/B est le déterminant majeur de la fonction transcriptionnelle différentielle des isoformes. Nous avons identifié plusieurs régulateurs transcriptionnels en tant que nouveaux partenaires d'interactions spécifiques pour certaines isoformes, tel le corépresseur IRF-2BP2. De façon intéressante, nous avons de plus montré pour la première fois une interaction potentielle de HNF4 α avec plusieurs larges complexes de remodelage de la chromatine dépendants de l'ATP comme les membres de la famille SWI/SNF, indépendamment des isoformes exprimées. Également, la redéfinition de la séquence réelle des isoformes HNF4 α 4, HNF4 α 5 et HNF4 α 6 nous a permis de montrer pour la première fois que l'activité transcriptionnelle de celles-ci est négligeable.

De façon globale, nous pouvons conclure que les 12 isoformes de HNF4 α sont des régulateurs de la transcription qui possèdent des activités et des partenaires d'interaction distincts, ce qui pourrait mener à des fonctions spécifiques pour ces différentes isoformes. L'analyse intégrative des données transcriptomiques et protéomiques générées au cours de ces travaux permettra dans le futur d'identifier des fonctions reliées à l'expression génique et qui sont spécifiques pour certaines isoformes de HNF4 α . Une question majeure, qui demeure encore sans réponse, concerne les mécanismes de régulation transcriptionnelle et d'épissage alternatif contrôlant l'expression spécifique des isoformes selon certains contextes. La complexité de cette question est d'autant plus importante considérant le haut

niveau de redondance fonctionnelle qui semble exister entre plusieurs isoformes. Certaines isoformes semblent souvent exprimées de façon concomitante et pourraient ainsi agir de façon synergique ou opposée. Les séquences extrêmement similaires des différentes isoformes limitent cependant le développement d'outils spécifiques à leur étude de manière endogène. Ainsi, au cours de ce projet, nous avons choisi une approche qui nous permettait d'étudier les isoformes de manière individuelle dans un contexte cellulaire. L'élaboration éventuelle d'un modèle de lignées stables permettant l'expression simultanée de plus d'une isoforme marquées de courtes étiquettes protéiques (myc, HA, V5, etc.) pourrait permettre d'analyser la relation fonctionnelle entre les différentes isoformes de HNF4 α .

Une grande quantité d'information demeure à être tirée des analyses effectuées lors de ce projet. L'identification de nouveaux corégulateurs potentiels de HNF4 α ouvre la porte sur de nouveaux mécanismes d'action spécifiques pour certaines isoformes et qui pourraient dépendre de ces interactions. La validation de la spécificité de ces interactions vis-à-vis les isoformes demeure encore à être effectuée. Dans l'ensemble, ce projet démontre que l'expression de plusieurs isoformes par HNF4 α représente un mécanisme important régulant la fonction transcriptionnelle générale de ce récepteur nucléaire. De façon intéressante, la similarité entre HNF4 α et les autres membres de la famille des récepteurs nucléaires ne s'arrête pas à leur structure hautement conservée. En effet, une grande quantité de récepteurs nucléaires sont connus pour être exprimés en plusieurs isoformes suite à l'utilisation de promoteurs alternatifs et de l'épissage alternatif. Les conclusions tirées suite à nos travaux pourraient ainsi servir de référence pour l'étude comparative d'autres récepteurs nucléaires exprimés en différentes isoformes.

REMERCIEMENTS

Je souhaite remercier en premier lieu le Pr Claude Asselin et le Pr Nicolas Gévry d'avoir généreusement accepté de prendre part à la correction de ce mémoire de maîtrise. Les commentaires reçus ainsi que leurs encouragements sont des mots précieux que je garderai longtemps en tête dans la poursuite de mes études. J'aimerais aussi remercier la Pre Marie-Josée Boucher pour les conversations stimulantes qu'elle a su apporter en étant membre de mon comité d'encadrement durant ma maîtrise.

J'aimerais remercier mes deux directeurs de recherche, le Pr François-Michel Boisvert et le Pr François Boudreau, pour m'avoir accueilli au sein de leur laboratoire lors de ma maîtrise. Je suis convaincu que je n'aurais pas pu choisir un meilleur environnement de travail pour réaliser ce projet. J'ai pu compter sur leur expertise respective en protéomique et en transcriptomique, et les nombreuses discussions scientifiques que nous avons partagées m'ont toujours fourni une motivation supplémentaire pour pousser plus loin mon projet de maîtrise. Je suis également reconnaissant pour la confiance qu'ils m'ont accordée, en me permettant de participer à différents congrès et de superviser des étudiants lors de mon projet et en tant qu'auxiliaire d'enseignement. J'apprécie particulièrement avoir pu bénéficier de leur mentorat, leur perspective et leur expérience qui m'ont grandement aidé dans mes réflexions pour la suite de mon parcours.

Je tiens aussi à dire merci à tous les membres des deux laboratoires Boisvert et Boudreau, qui ont été des collègues et des amis très importants au cours de ma maîtrise. Les souvenirs des nombreuses fêtes, congrès et fameux costumes d'Halloween vont demeurer longtemps dans ma mémoire. Merci à Dominique Lévesque, qui a été la première personne à me superviser dans le laboratoire Boisvert lors de mon tout premier stage en recherche. Son aide et ses connaissances ont été primordiales à la réalisation de mon projet de maîtrise, surtout en lien avec le clonage moléculaire et la spectrométrie de masse. Je remercie aussi Dominique Jean pour son travail en lien avec la plateforme de protéomique de l'Université de Sherbrooke. J'aimerais remercier Christine Jones, qui a toujours été là en tant qu'assistante de recherche pour m'accueillir dans le laboratoire Boudreau. Je suis heureux de connaître une personne aussi dévouée et à l'écoute, autant au travail qu'en dehors du laboratoire. Également, je tiens absolument à souligner le rôle essentiel que

Jean-Philippe Babeu a joué dans mon cheminement. Sa générosité, son attitude positive et sa persévérance sont des qualités que j'associerai maintenant toujours à lui et que je tente de suivre de mon mieux. Je réalise ma chance d'avoir pu compter sur un mentor comme lui durant ma maîtrise.

J'aimerais remercier Marie-Line Dubois et Anaïs Chauvin, qui ont su mettre de la vie dans mon travail au laboratoire de manière quotidienne. Elles ont été comme des grandes sœurs pour moi, toujours là pour prêter un coup de main ou une oreille, ou encore pour une bière de fin de journée ensemble dans le bureau. Je ne suis pas inquiet que nous aurons la chance de se revoir souvent malgré mon départ. Je remercie aussi les personnes qui se jointes à mon projet au cours de ma maîtrise, notamment ma stagiaire Émilie Jolibois, ainsi que Joël Simoneau et la Pre Michelle Scott pour leur aide précieuse dans la partie séquençage de l'ARN du projet. J'aimerais également remercier Maryse Adam pour son aide avec le protocole de retard sur gel, et avec qui j'ai eu la chance de travailler en tant qu'auxiliaire de recherche. Cette expérience a été très stimulante et formative, et ce fut un réel plaisir de la côtoyer durant mon parcours à Sherbrooke. Finalement, je remercie ma famille et mes amis pour leur support et leurs encouragements en dehors du laboratoire, qui ont contribué à leur façon à mon cheminement.

RÉFÉRENCES

- Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., et Walter, P. (2015) Control of gene expression. *Molecular biology of the cell*. New York : Garland Science (6^e éd.), 407-38.
- Andrews, S. (2016). FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>, page consultée le 20 novembre 2018.
- Babeu, J. P. (2016) *Les isoformes P1 et P2 du récepteur nucléaire HNF4a ont des fonctions différentes dans le cancer colorectal* (Dissertation doctorale). Université de Sherbrooke, Sherbrooke, Canada.
- Babeu, J. P., Jones, C., Geha, S., Carrier, J. C., Boudreau, F. (2018) P1 promoter-driven HNF4a isoforms are specifically repressed by β -catenin signaling in colorectal cancer cells. *J Cell Sci*, 131(13). pii: jcs214734.
- Bain, D. L., Heneghan, A. F., Connaghan-Jones, K. D., et Miura, M. T. (2007) Nuclear receptor structure: implications for function. *Annu Rev Physiol*, 69: 201-20.
- Barrero, M. J., et Malik, S. (2006) Two functional modes of a nuclear receptor-recruited arginine methyltransferase in transcriptional activation. *Mol Cell*, 24(2): 233-43.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., Wilson, C. J., Lehár, J., Kryukov, G. V., ... Garraway, L. A. (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391): 603-7.
- Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., et Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4): 823-37.
- Basta, J., et Rauchman, M. (2015) The nucleosome remodeling and deacetylase complex in development and disease. *Transl Res*, 165(1): 36-47.
- Benecke, A., Chambon, P., et Gronemeyer, H. (2000) Synergy between estrogen receptor alpha activation functions AF1 and AF2 mediated by transcription intermediary factor TIF2. *EMBO Rep*, 1(2): 151-7.
- Boj, S. F., Parrizas, M., Maestro, M. A., et Ferrer, J. (2001) A transcription factor regulatory circuit in differentiated pancreatic cells. *Proc Natl Acad Sci U S A*, 98(25) : 14481-6.
- Bolger, A. M., Lohse, M., et Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.
- Bommer, M., Benecke, A., Gronemeyer, H., et Rochette-Egly, C. (2002) TIF2 mediates the synergy between RARalpha 1 activation functions AF-1 and AF-2. *J Biol Chem*, 277(40): 37961-6.

- Boyd, M., Bressendorff, S., Møller, J., Olsen, J., et Troelsen, J. T. (2009) Mapping of HNF4alpha target genes in intestinal epithelial cells. *BMC Gastroenterol*, 9:68.
- Bray, N. L., Pimentel, H., Melsted, P., et Pachter, L. (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*, 34(5): 525-7.
- Briançon, N., Bailly, A., Clotman, F., Jacquemin, P., Lemaigre, F. P., et Weiss, M. C. (2004) Expression of the alpha7 isoform of hepatocyte nuclear factor (HNF) 4 is activated by HNF6/OC-2 and HNF1 and repressed by HNF4alpha1 in the liver. *J Biol Chem*, 279(32): 33398-408.
- Briançon, N., et Weiss, M. C. (2006) In vivo role of the HNF4alpha AF-1 activation domain revealed by exon swapping. *EMBO J*, 25(6): 1253-62.
- Carneiro, F. R., Ramalho-Oliveira, R., Mognol, G. P., et Viola, J. P. (2011) Interferon regulatory factor 2 binding protein 2 is a new NFAT1 partner and represses its transcriptional activity. *Mol Cell Biol*, 31(14): 2889-901.
- Cattin, A. L., Le Beyec, J., Barreau, F., Saint-Just, S., Houllier, A., Gonzalez, F. J., Robine, S., Pinçon-Raymond, M., Cardot, P., ... Ribeiro, A. (2009) Hepatocyte nuclear factor 4alpha, a key factor for homeostasis, cell architecture, and barrier function of the adult intestinal epithelium. *Mol Cell Biol*, 29(23): 6294-308.
- Chandra, V., Huang, P., Potluri, N., Wu, D., Kim, Y., et Rastinejad, F. (2013) Multidomain integration in the structure of the HNF-4a nuclear receptor complex. *Nature*, 495(7441): 394-8.
- Chartier, F. L., Bossu, J. P., Laudet, V., Fruchart, J. C., et Laine, B. (1994) Cloning and sequencing of cDNAs encoding the human hepatocyte nuclear factor 4 indicate the presence of two isoforms in human liver. *Gene*, 147(2): 269-72.
- Chauvin, A., Wang, C. S., Geha, S., Garde-Granger, P., Mathieu, A. A., Lacasse, V., Boisvert, F. M. (2018) The response to neoadjuvant chemoradiotherapy with 5-fluorouracil in locally advanced rectal cancer patients: a predictive proteomic signature. *Clin Proteomics*, 15:16.
- Chellappa, K., Jankova, L., Schnabl, J. M., Pan, S., Brelivet, Y., Fung, C. L., Chan, C., Dent, O. F., Clarke, S. J., ... Sladek, F. M. (2012) Src tyrosine kinase phosphorylation of nuclear receptor HNF4a correlates with isoform-specific loss of HNF4a in human colon cancer. *Proc Natl Acad Sci U S A*, 109(7): 2302-7.
- Chellappa, K., Deol, P., Evans, J. R., Vuong, L. M., Chen, G., Briançon, N., Bolotin, E., Lytle, C., Nair, M. G., et Sladek, F. M. (2016) Opposing roles of nuclear receptor HNF4a isoforms in colitis and colitis-associated colon cancer. *Elife*, 5. pii: e10903.
- Chen, W. S., Manova, K., Weinstein, D. C., Duncan, S. A., Plump, A. S., Prezioso, V. R., Bachvarova, R. F., et Darnell, J. E. Jr. (1994) Disruption of the HNF-4 gene, expressed in visceral endoderm, leads to cell death in embryonic ectoderm and impaired gastrulation of mouse embryos. *Genes Dev*, 8(20): 2466-77.

Childs, K. S., et Goodbourn, S. (2003) Identification of novel co-repressor molecules for Interferon Regulatory Factor-2. *Nucleic Acids Res*, 31(12): 3016-26.

Chou, W. C., Prokova, V., Shiraishi, K., Valcourt, U., Moustakas, A., Hadzopoulou-Cladaras, M., Zannis, V. I., et Kardassis, D. (2003) Mechanism of a transcriptional cross talk between transforming growth factor-beta-regulated Smad3 and Smad4 proteins and orphan nuclear receptor hepatocyte nuclear factor-4. *Mol Biol Cell*, 14(3): 1279-94.

Costa, R. H., Grayson, D. R., et Darnell, J. E. Jr. (1989) Multiple hepatocyte-enriched nuclear factors function in the regulation of transthyretin and alpha 1-antitrypsin genes. *Mol Cell Biol*, 9(4): 1415-25.

Cox, J., et Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*, 26(12): 1367-72.

Dasgupta, S., Lonard, D. M., et O'Malley, B. W. (2014) Nuclear receptor coactivators: master regulators of human health and disease. *Annu Rev Med*, 65: 279-92.

DeLaForest, A., Nagaoka, M., Si-Tayeb, K., Noto, F. K., Konopka, G., Battle, M. A., et Duncan, S. A. (2011) HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. *Development*, 138(19): 4143-53.

Dell, H., et Hadzopoulou-Cladaras, M. (1999) CREB-binding protein is a transcriptional coactivator for hepatocyte nuclear factor-4 and enhances apolipoprotein gene expression. *J Biol Chem*, 274(13): 9013-21.

Dhe-Paganon, S., Duda, K., Iwamoto, M., Chi, Y. I., et Shoelson, S. E. (2002) Crystal structure of the HNF4 alpha ligand binding domain in complex with endogenous fatty acid ligand. *J Biol Chem*, 277(41): 37973-6.

Drewes, T., Senkel, S., Holewa, B., et Ryffel, G. U. (1996) Human hepatocyte nuclear factor 4 isoforms are encoded by distinct and differentially expressed genes. *Mol Cell Biol*, 16(3): 925-31.

Duncan, S. A., Navas, M. A., Dufort, D., Rossant, J., et Stoffel, M. (1998) Regulation of a transcription factor network required for differentiation and metabolism. *Science*, 281(5377): 692-5.

Eeckhoute, J., Moerman, E., Bouckennooghe, T., Lukoviak, B., Pattou, F., Formstecher, P., Kerr-Conte, J., Vandewalle, B., et Laine, B. (2003) Hepatocyte nuclear factor 4 alpha isoforms originated from the P1 promoter are expressed in human pancreatic beta-cells and exhibit stronger transcriptional potentials than P2 promoter-driven isoforms. *Endocrinology*, 144(5): 1686-94.

Eeckhoute, J., Formstecher, P., et Laine, B. (2004) Hepatocyte nuclear factor 4alpha enhances the hepatocyte nuclear factor 1alpha-mediated activation of transcription. *Nucleic Acids Res*, 32(8): 2586-93.

Erdmann, S., Senkel, S., Arndt, T., Lucas, B., Lausen, J., Klein-Hitpass, L., Ryffel, G. U., et Thomas, H. (2007) Tissue-specific transcription factor HNF4alpha inhibits cell proliferation and induces apoptosis in the pancreatic INS-1 beta-cell line. *Biol Chem*, 388(1): 91-106.

Fang, B., Mane-Padros, D., Bolotin, E., Jiang, T., et Sladek, F. M. (2012) Identification of a binding motif specific to HNF4 by comparative analysis of multiple nuclear receptors. *Nucleic Acids Res*, 40(12): 5343-56.

Ferrer, J. (2002) A genetic switch in pancreatic beta-cells: implications for differentiation and haploinsufficiency. *Diabetes*, 51(8): 2355-62.

Formstecher, P., et Laine, B. (1999) The activity of the activation function 2 of the human hepatocyte nuclear factor 4 (HNF-4alpha) is differently modulated by F domains from various origins. *Biochem J*, 340(Pt 1): 161-9.

Furuta, H., Iwasaki, N., Oda, N., Hinokio, Y., Horikawa, Y., Yamagata, K., Yano, N., Sugahiro, J., Ogata, M., ... Bell, G. I. (1997) Organization and partial sequence of the hepatocyte nuclear factor-4 alpha/MODY1 gene and identification of a missense mutation, R127W, in a Japanese family with MODY. *Diabetes*, 46(10): 1652-7.

Garrison, W. D., Battle, M. A., Yang, C., Kaestner, K. H., Sladek, F. M., et Duncan, S. A. (2006) Hepatocyte nuclear factor 4alpha is essential for embryonic development of the mouse colon. *Gastroenterology*, 130(4): 1207-20.

Germain, P., Staels, B., Dacquet, C., Spedding, M., et Laudet, V. (2006) Overview of nomenclature of nuclear receptors. *Pharmacol Rev*, 58(4): 685-704.

Giangrande, P. H., et McDonnell, D. P. (1999) The A and B isoforms of the human progesterone receptor: two functionally different transcription factors encoded by a single gene. *Recent Prog Horm Res*, 54: 291-313.

Giguère, V., Tini, M., Flock, G., Ong, E., Evans, R. M., et Otulakowski, G. (1994) Isoform-specific amino-terminal domains dictate DNA-binding properties of ROR alpha, a novel family of orphan hormone nuclear receptors. *Genes Dev*, 8(5): 538-53.

Giguère, V. (1999) Orphan nuclear receptors: from gene to function. *Endocr Rev*, 20(5): 689-725.

Gould, J. (2018). Morpheus. <https://software.broadinstitute.org/morpheus>, page consultée le 8 novembre 2018.

Green, V. J., Kokkotou, E., et Ladas, J. A. (1998) Critical structural elements and multitarget protein interactions of the transcriptional activator AF-1 of hepatocyte nuclear factor 4. *J Biol Chem*, 273(45): 29950-7.

Gronemeyer, H., Gustafsson, J. A., et Laudet, V. (2004) Principles for modulation of the nuclear receptor superfamily. *Nat Rev Drug Discov*, 3(11): 950-64.

Gu, B., et Lee, M. G. (2013) Histone H3 lysine 4 methyltransferases and demethylases in self-renewal and differentiation of stem cells. *Cell Biosci*, 3(1): 39.

Gupta, R. K., Vatamaniuk, M. Z., Lee, C. S., Flaschen, R. C., Fulmer, J. T., Matschinsky, F. M., Duncan, S. A., et Kaestner, K. H. (2005) The MODY1 gene HNF-4alpha regulates selected genes involved in insulin secretion. *J Clin Invest*, 115(4): 1006-15.

Gupta, R. K., Gao, N., Gorski, R. K., White, P., Hardy, O. T., Rafiq, K., Brestelli, J. E., Chen, G., Stoeckert, C. J. Jr., et Kaestner, K. H. (2007) Expansion of adult beta-cell mass in response to increased metabolic demand is dependent on HNF-4alpha. *Genes Dev*, 21(7): 756-69.

Hadzopoulou-Cladaras, M., Kistanova, E., Evagelopoulou, C., Zeng, S., Cladaras, C., et Ladias, J. A. (1997) Functional domains of the nuclear receptor hepatocyte nuclear factor 4. *J Biol Chem*, 272(1): 539-50.

Hansen, S. K., Párrizas, M., Jensen, M. L., Pruhova, S., Ek, J., Boj, S. F., Johansen, A., Maestro, M. A., Rivera, F., ... Hansen, T. (2002) Genetic evidence that HNF-1alpha-dependent transcriptional control of HNF-4alpha is essential for human pancreatic beta cell function. *J Clin Invest*, 110(6): 827-33.

Harries, L. W., Locke, J. M., Shields, B., Hanley, N. A., Hanley, K. P., Steele, A., Njølstad, P. R., Ellard, S., et Hattersley, A. T. (2008) The diabetic phenotype in HNF4A mutation carriers is moderated by the expression of HNF4A isoforms from the P1 promoter during fetal development. *Diabetes*, 57(6): 1745-52.

Hatzis, P., et Talianidis, I. (2001) Regulatory mechanisms controlling human hepatocyte nuclear factor 4alpha gene expression. *Mol Cell Biol*, 21(21):7320-30.

Hayhurst, G. P., Lee, Y. H., Lambert, G., Ward, J. M., et Gonzalez, F. J. (2001) Hepatocyte nuclear factor 4alpha (nuclear receptor 2A1) is essential for maintenance of hepatic gene expression and lipid homeostasis. *Mol Cell Biol*, 21(4): 1393-403.

Heery, D. M., Kalkhoven, E., Hoare, S., et Parker, M. G. (1997) A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature*, 387(6634): 733-6.

Hong, Y. H., Varanasi, U. S., Yang, W., et Leff, T. (2003) AMP-activated protein kinase regulates HNF4alpha transcriptional activity by inhibiting dimer formation and decreasing protein stability. *J Biol Chem*, 278(30): 27495-501.

Hota, S. K., et Bruneau, B. G. (2016) ATP-dependent chromatin remodeling during mammalian development. *Development*, 143(16): 2882-97.

Hsia, E. Y., Goodson, M. L., Zou, J. X., Privalsky, M. L., et Chen, H. W. (2010) Nuclear receptor coregulators as a new paradigm for therapeutic targeting. *Adv Drug Deliv Rev*, 62(13): 1227-37.

Huang, J., Karakucuk, V., Levitsky, L. L., et Rhoads, D. B. (2008) Expression of HNF4alpha variants in pancreatic islets and Ins-1 beta cells. *Diabetes Metab Res Rev*, 24(7): 533-43.

Huang, J., Levitsky, L. L., et Rhoads, D. B. (2009) Novel P2 promoter-derived HNF4alpha isoforms with different N-terminus generated by alternate exon insertion. *Exp Cell Res*, 315(7): 1200-11.

Huang, P., Chandra, V., et Rastinejad, F. (2010) Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. *Annu Rev Physiol*, 72: 247-72.

Inoue, Y., Yu, A. M., Yim, S. H., Ma, X., Krausz, K. W., Inoue, J., Xiang, C. C., Brownstein, M. J., Eggertsen, G., ... Gonzalez, F. J. (2006) Regulation of bile acid biosynthesis by hepatocyte nuclear factor 4alpha. *J Lipid Res*, 47(1): 215-27.

Iyemere, V. P., Davies, N. H., et Brownlee, G. G. (1998) The activation function 2 domain of hepatic nuclear factor 4 is regulated by a short C-terminal proline-rich repressor domain. *Nucleic Acids Res*, 26(9): 2098-104.

Jacquet, K., Fradet-Turcotte, A., Avvakumov, N., Lambert, J. P., Roques, C., Pandita, R. K., Paquet, E., Herst, P., Gingras, A. C., ... Côté, J. (2016) The TIP60 Complex Regulates Bivalent Chromatin Recognition by 53BP1 through Direct H4K20me Binding and H2AK15 Acetylation. *Mol Cell*, 62(3): 409-21.

Jiang, G., Nepomuceno, L., Hopkins, K., et Sladek, F. M. (1995) Exclusive homodimerization of the orphan receptor hepatocyte nuclear factor 4 defines a new subclass of nuclear receptors. *Mol Cell Biol*, 15(9): 5131-43.

Jiang, G., et Sladek, F. M. (1997) The DNA binding domain of hepatocyte nuclear factor 4 mediates cooperative, specific binding to DNA and heterodimerization with the retinoid X receptor alpha. *J Biol Chem*, 272(2): 1218-25.

Jover, R., Bort, R., Gómez-Lechón, M. J., et Castell, J. V. (2001) Cytochrome P450 regulation by hepatocyte nuclear factor 4 in human hepatocytes: a study using adenovirus-mediated antisense targeting. *Hepatology*, 33(3): 668-75.

Kastner, P., Krust, A., Turcotte, B., Stropp, U., Tora, L., Gronemeyer, H., et Chambon, P. (1990) Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. *EMBO J*, 9(5): 1603-14.

Khorasanizadeh, S., et Rastinejad, F. (2001) Nuclear-receptor interactions on DNA-response elements. *Trends Biochem Sci*, 26(6): 384-90.

Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K., et Roux, K. J. (2016) An improved smaller biotin ligase for BioID proximity labeling. *Mol Cell Biol*, 27(8): 1188-96.

Kimura, M. (2008) IRF2-binding protein-1 is a JDP2 ubiquitin ligase and an inhibitor of ATF2-dependent transcription. *FEBS Lett*, 582(19): 2833-7.

Koeppel, M., van Heeringen, S. J., Smeenk, L., Navis, A. C., Janssen-Megens, E. M., et Lohrum, M. (2009) The novel p53 target gene IRF2BP2 participates in cell survival during the p53 stress response. *Nucleic Acids Res*, 37(2): 322-35.

- Krasnov, A. N., Mazina, M. Y., Nikolenko, J. V., et Vorobyeva, N. E. (2016) On the way of revealing coactivator complexes cross-talk during transcriptional activation. *Cell Biosci*, 6:15.
- Kritis, A. A., Argyrokastritis, A., Moschonas, N. K., Power, S., Katrakili, N., Zannis, V. I., Cereghini, S., et Talianidis, I. (1996) Isolation and characterization of a third isoform of human hepatocyte nuclear factor 4. *Gene*, 173(2): 275-80.
- Ktistaki, E., et Talianidis, I. (1997) Chicken ovalbumin upstream promoter transcription factors act as auxiliary cofactors for hepatocyte nuclear factor 4 and enhance hepatic gene expression. *Mol Cell Biol*, 17(5): 2790-7.
- Ktistaki, E., et Talianidis, I. (1997) Modulation of hepatic gene expression by hepatocyte nuclear factor 1. *Science*, 277(5322): 109-12.
- Kyrmizi, I., Hatzis, P., Katrakili, N., Tronche, F., Gonzalez, F. J., et Talianidis, I. (2006) Plasticity and expanding complexity of the hepatic transcription factor network during liver development. *Genes Dev*, 20(16): 2293-305.
- Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., ... Weirauch, M. T. (2018) The human transcription factors. *Cell*, 172(4): 650-65.
- Lau, H. H., Ng, N. H. J., Loo, L. S. W., Jasmen, J. B., et Teo, A. K. K. (2018) The molecular functions of hepatocyte nuclear factors - In and beyond the liver. *J Hepatol*, 68(5): 1033-48.
- Laudet, V. (2006) Nuclear receptor genes. *Encyclopedia of Life Sciences*, DOI: 10.1038/npg.els.0006154.
- Leclerc, I., Lenzner, C., Gourdon, L., Vaulont, S., Kahn, A., et Viollet, B. (2001) Hepatocyte nuclear factor-4 α involved in type 1 maturity-onset diabetes of the young is a novel target of AMP-activated protein kinase. *Diabetes*, 50(7): 1515-21.
- Lee, Y. K., Dell, H., Dowhan, D. H., Hadzopoulou-Cladaras, M., et Moore, D. D. (2000) The orphan nuclear receptor SHP inhibits hepatocyte nuclear factor 4 and retinoid X receptor transactivation: two mechanisms for repression. *Mol Cell Biol*, 20(1): 187-95.
- Lee, S., et Privalsky, M. L. (2005) Heterodimers of retinoic acid receptors and thyroid hormone receptors display unique combinatorial regulatory properties. *Mol Endocrinol*, 19(4): 863-78.
- Li, J., Ning, G., et Duncan, S. A. (2000) Mammalian hepatocyte differentiation requires the transcription factor HNF-4 α . *Genes Dev*, 14(4): 464-74.
- Lonard, D. M., et O'Malley, B. W. (2012) Nuclear receptor coregulators: modulators of pathology and therapeutic targets. *Nat Rev Endocrinol*, 8(10): 598-604.
- Love, M., Huber, W., et Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12): 550.

- Lu, H. (2016) Crosstalk of HNF4 α with extracellular and intracellular signaling pathways in the regulation of hepatic metabolism of drugs and lipids. *Acta Pharm Sin B*, 6(5): 393-408.
- Lucas, B., Grigo, K., Erdmann, S., Lausen, J., Klein-Hitpass, L., et Ryffel, G. U. (2005) HNF4 α reduces proliferation of kidney cells and affects genes deregulated in renal cell carcinoma. *Oncogene*, 24(42): 6418-31.
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., et Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648): 251-60.
- Luscombe, N. M., Austin, S. E., Berman, H. M., et Thornton, J. M. (2000) An overview of the structures of protein-DNA complexes. *Genome Biol*, 1(1): REVIEWS001.
- Lussier, C. R., Babeu, J. P., Auclair, B. A., Perreault, N., et Boudreau, F. (2008) Hepatocyte nuclear factor-4 α promotes differentiation of intestinal epithelial cells in a coculture system. *Am J Physiol Gastrointest Liver Physiol*, 294(2): G-418-28.
- Maeda, Y., Rachez, C., Hawel, L., Byus, C. V., Freedman, L. P., et Sladek, F. M. (2002) Polyamines modulate the interaction between nuclear receptors and vitamin D receptor-interacting protein 205. *Mol Endocrinol*, 16(7): 1502-10.
- Malik, S., Guermah, M., et Roeder, R. G. (1998) A dynamic model for PC4 coactivator function in RNA polymerase II transcription. *Proc Natl Acad Sci U S A*, 95(5): 2192-7.
- Malik, S., Wallberg, A. E., Kang, Y. K., et Roeder, R. G. (2002) TRAP/SMCC/mediator-dependent transcriptional activation from DNA and chromatin templates by orphan nuclear receptor hepatocyte nuclear factor 4. *Mol Cell Biol*, 22(15): 5626-37.
- Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schütz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., ... Evans, R. M. (1995) The nuclear receptor superfamily: the second decade. *Cell*, 83(6): 835-9.
- Margueron, R., et Reinberg, D. (2011) The Polycomb complex PRC2 and its mark in life. *Nature*, 469(7330): 343-9.
- Martínez-Jiménez, C. P., Gómez-Lechón, M. J., Castell, J. V., et Jover, R. (2006) Underexpressed coactivators PGC1 α and SRC1 impair hepatocyte nuclear factor 4 α function and promote dedifferentiation in human hepatoma cells. *J Biol Chem*, 281(40): 29840-9.
- Maston, G. A., Evans, S. K., et Green, M. R. (2006) Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet*, 7: 29-59.
- Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., Lewis, S., et Thomas, P. D. (2010) PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res*, 38: D204-10.

- Mietus-Snyder, M., Sladek, F. M., Ginsburg, G. S., Kuo, C. F., Ladas, J. A., Darnell, J. E. Jr., et Karathanasis, S. K. (1992) Antagonism between apolipoprotein AI regulatory protein 1, Ear3/COUP-TF, and hepatocyte nuclear factor 4 modulates apolipoprotein CIII gene expression in liver and intestinal cells. *Mol Cell Biol*, 12(4): 1708-18.
- Millard, C. J., Watson, P. J., Fairall, L., et Schwabe, J. W. (2013) An evolving understanding of nuclear receptor coregulator proteins. *J Mol Endocrinol*, 51(3): T23-36.
- Näär, A. M., Lemon, B. D., et Tjian, R. (2001) Transcriptional coactivator complexes. *Annu Rev Biochem*, 70: 475-501.
- Nakhei, H., Lingott, A., Lemm, I., et Ryffel, G. U. (1998) An alternative splice variant of the tissue specific transcription factor HNF4alpha predominates in undifferentiated murine cell types. *Nucleic Acids Res*, 26(2): 497-504.
- Nikolaidou-Neokosmidou, V., Zannis, V. I., et Kardassis, D. (2006) Inhibition of hepatocyte nuclear factor 4 transcriptional activity by the nuclear factor kappaB pathway. *Biochem J*, 398(3): 439-50.
- Nuclear Receptors Nomenclature Committee. (1999) A unified nomenclature system for the nuclear receptor superfamily. *Cell*, 97(2): 161-3.
- Nyirenda, M. J., Dean, S., Lyons, V., Chapman, K. E., Seckl, J. R. (2006) Prenatal programming of hepatocyte nuclear factor 4alpha in the rat: A key mechanism in the 'foetal origins of hyperglycaemia'? *Diabetologia*, 49(6): 1412-20.
- Oakley, R. H., Sar, M., et Cidlowski, J. A. (1996) The human glucocorticoid receptor beta isoform. Expression, biochemical properties, and putative function. *J Biol Chem*, 271(16): 9550-9.
- Odom, D. T., Zizlsperger, N., Gordon, D. B., Bell, G. W., Rinaldi, N. J., Murray, H. L., Volkert, T. L., Schreiber, J., Rolfe, P. A., ... Young, R. A. (2004) Control of pancreas and liver gene expression by HNF transcription factors. *Science*, 303(5662): 1378-81.
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., ... Pruitt, K.D. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*, 44(D1): D733-45.
- Ong, C. T., et Corces, V. G. (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet*, 12(4): 283-93.
- Pascussi, J. M., Robert, A., Moreau, A., Ramos, J., Bioulac-Sage, P., Navarro, F., Blanc, P., Assenat, E., Maurel, P., et Vilarem, M. J. (2007) Differential regulation of constitutive androstane receptor expression by hepatocyte nuclear factor4alpha isoforms. *Hepatology*, 45(5): 1146-53.
- Patel, S. R., et Skafar, D. F. (2015) Modulation of nuclear receptor activity by the F domain. *Mol Cell Endocrinol*, 418: 298-305.

- Pawlak, M., Lefebvre, P., et Staels, B. (2012) General molecular biology and architecture of nuclear receptors. *Curr Top Med Chem*, 12(6): 486-504.
- Pollard, Earnshaw, Lippincott-Schwartz, et Johnson. (2016) Gene expression. *Cell biology*. Philadelphia : Elsevier (3^e éd.), 165-88.
- Poss, Z. C., Ebmeier, C. C., et Taatjes, D. J. (2013) The Mediator complex and transcription regulation. *Crit Rev Biochem Mol Biol*, 48(6): 575-608.
- Rana, R., Surapureddi, S., Kam, W., Ferguson, S., et Goldstein, J. A. (2011) Med25 is required for RNA polymerase II recruitment to specific promoters, thus regulating xenobiotic and lipid metabolism in human liver. *Mol Cell Biol*, 31(3): 466-81.
- Reiter, F., Wienerroither, S., et Stark, A. (2017) Combinatorial function of transcription factors and cofactors. *Curr Opin Genet Dev*, 43: 73-81.
- Robinson-Rechavi, M., Escriva, G. H., et Laudet, V. (2003) The nuclear receptor superfamily. *J Cell Sci*, 116(4): 585-6.
- Rosenfeld, M. G., Lunyak, V. V., et Glass, C. K. (2006) Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes Dev*, 20(11): 1405-28.
- Roux, K. J., Kim, D. I., Raida, M., et Burke, B. (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol*, 196(6): 801-10.
- Ruse, M. D. Jr., Privalsky, M. L., et Sladek, F. M. (2002) Competitive cofactor recruitment by orphan receptor hepatocyte nuclear factor 4alpha1: modulation by the F domain. *Mol Cell Biol*, 22(6): 1626-38.
- Sainsbury, S., Bernecky, C., et Cramer, P. (2015) Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol*, 16(3): 129-43.
- Sasaki, S., Urabe, M., Maeda, T., Suzuki, J., Irie, R., Suzuki, M., Tomaru, Y., Sakaguchi, M., Gonzalez, F. J., et Inoue, Y. (2018) Induction of hepatic metabolic functions by a novel variant of hepatocyte nuclear factor 4γ. *Mol Cell Biol*, pii: MCB.00213-18.
- Schwartz, B., Algamas-Dimantov, A., Hertz, R., Nataf, J., Kerman, A., Peri, I., et Bar-Tana, J. (2009) Inhibition of colorectal cancer by targeting hepatocyte nuclear factor-4alpha. *Int J Cancer*, 124(5): 1081-9.
- Sever, R., et Glass, C. K. (2013) Signaling by nuclear receptors. *Cold Spring Harb Perspect Biol*, 5(3): a016709.
- Sims, R. J. III, Belotserkovskaya, R., et Reinberg, D. (2004) Elongation by RNA polymerase II: the short and long of it. *Genes Dev*, 18(20): 2437-68.

Sladek, F. M., Zhong, W. M., Lai, E., et Darnell, J. E. Jr. (1990) Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Genes Dev*, 4(12B): 2353-65.

Sladek, F. M., Ruse, M. D. Jr., Nepomuceno, L., Huang, S. M., et Stallcup, M. R. (1999) Modulation of transcriptional activation and coactivator interaction by a splicing variation in the F domain of nuclear receptor hepatocyte nuclear factor 4alpha1. *Mol Cell Biol*, 19(10): 6509-22.

Soutoglou, E., Katrakili, N., et Talianidis, I. (2000) Acetylation regulates transcription factor activity at multiple levels. *Mol Cell*, 5(4): 745-51.

Stewart, M. D., et Wong, J. (2009) Nuclear receptor repression: regulatory mechanisms and physiological implications. *Prog Mol Biol Transl Sci*, 87: 235-59.

Stoffel, M., et Duncan, S. A. (1997) The maturity-onset diabetes of the young (MODY1) transcription factor HNF4alpha regulates expression of genes required for glucose transport and metabolism. *Proc Natl Acad Sci U S A*, 94(24): 13209-14.

Sun, K., Montana, V., Chellappa, K., Brelivet, Y., Moras, D., Maeda, Y., Parpura, V., Paschal, B. M., et Sladek, B. M. (2007) Phosphorylation of a conserved serine in the deoxyribonucleic acid binding domain of nuclear receptors alters intracellular localization. *Mol Endocrinol*, 21(6): 1297-311.

Tanaka, T., Jiang, S., Hotta, H., Takano, K., Iwanari, H., Sumi, K., Daigo, K., Ohashi, R., Sugai, M., ... Kodama, T. (2006). Dysregulated expression of P1 and P2 promoter-driven hepatocyte nuclear factor-4alpha in the pathogenesis of human cancer. *J Pathol*, 208(5): 662-72.

Thomas, H., Jaschowitz, K., Bulman, M., Frayling, T. M., Mitchell, S. M., Roosen, S., Lingott-Frieg, A., Tack, C. J., Ellard, S., ... et Hattersley, A. T. (2001) A distant upstream promoter of the HNF-4alpha gene connects the transcription factors involved in maturity-onset diabetes of the young. *Hum Mol Genet*, 10(19): 2089-97.

Torres-Padilla, M. E., Fougère-Deschatrette, C., et Weiss, M. C. (2001) Expression of HNF4alpha isoforms in mouse liver development is regulated by sequential promoter usage and constitutive 3' end splicing. *Mech Dev*, 109(2): 183-93.

Torres-Padilla, M. E., Sladek, F. M., et Weiss, M. C. (2002) Developmentally regulated N-terminal variants of the nuclear receptor hepatocyte nuclear factor 4alpha mediate multiple interactions through coactivator and corepressor-histone deacetylase complexes. *J Biol Chem*, 277(47): 44677-87.

Urnov, F. D., Wolffe, A. P., et Guschin, D. (2001) Molecular mechanisms of corepressor function. Dans M. L. Privalsky (Éd). *Transcriptional corepressors: mediators of eukaryotic gene repression*. New York : Springer (1^e éd.), 1-34.

- Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., et Luscombe, N.M. (2009) A census of human transcription factors: function, expression and evolution. *Nat Rev Genet*, 10(4): 252-63.
- Viollet, B., Kahn, A., et Raymondjean, M. (1997) Protein kinase A-dependent phosphorylation modulates DNA-binding activity of hepatocyte nuclear factor 4. *Mol Cell Biol*, 17(8): 4208-19.
- Vuong, L. M., Chellappa, K., Dhahbi, J. M., Deans, J. R., Fang, B., Bolotin, E., Titova, N. V., Hoverter, N. P., Spindler, S. R., ... Sladek, F. M. (2015) Differential Effects of Hepatocyte Nuclear Factor 4 α Isoforms on Tumor Growth and T-Cell Factor 4/AP-1 Interactions in Human Colorectal Cancer Cells. *Mol Cell Biol*, 35(20): 3471-90.
- Wang, J. C., Stafford, J. M., et Granner, D. K. (1998) SRC-1 and GRIP1 coactivate transcription with hepatocyte nuclear factor 4. *J Biol Chem*, 273(47): 30847-50.
- Wang, H., Maechler, P., Antinozzi, P. A., Hagenfeldt, K. A., et Wollheim, C. B. (2000) Hepatocyte nuclear factor 4 α regulates the expression of pancreatic beta -cell genes implicated in glucose metabolism and nutrient-induced insulin secretion. *J Biol Chem*, 275(46): 35953-9.
- Wärnmark, A., Treuter, E., Wright, A. P., et Gustafsson, J. A. (2003) Activation functions 1 and 2 of nuclear receptors: molecular strategies for transcriptional activation. *Mol Endocrinol*, 17(10): 1901-9.
- Watson, P. J., Fairall, L., et Schwabe, J. W. (2012) Nuclear hormone receptor co-repressors: structure and function. *Mol Cell Endocrinol*, 348(2): 440-9.
- Wei, L. L., Hawkins, P., Baker, C., Norris, B., Sheridan, P. L., et Quinn, P. G. (1996) An amino-terminal truncated progesterone receptor isoform, PRC, enhances progestin-induced transcriptional activity. *Mol Endocrinol*, 10(11): 1379-87.
- Woodcock, C. L., et Ghosh, R. P. (2010) Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol*, 2(5): a000596.
- Xie, Y. B., Nedumaran, B., et Choi, H. S. (2009) Molecular characterization of SMILE as a novel corepressor of nuclear receptors. *Nucleic Acids Res*, 37(12): 4100-15.
- Xu, J., et Li, Q. (2003) Review of the in vivo functions of the p160 steroid receptor coactivator family. *Mol Endocrinol*, 17(9): 1681-92.
- Yamamoto, T., Shimano, H., Nakagawa, Y., Ide, T., Yahagi, N., Matsuzaka, T., Nakakuki, M., Takahashi, A., Suzuki, H., ... Yamada, N. (2004) SREBP-1 interacts with hepatocyte nuclear factor-4 α and interferes with PGC-1 recruitment to suppress hepatic gluconeogenic genes. *J Biol Chem*, 279(13): 12027-35.
- Yang, Y. Z., Burgos-Trinidad, M., Wu, Y., et Koenig, R. J. (1996) Thyroid hormone receptor variant α 2. Role of the ninth heptad in dna binding, heterodimerization with retinoid X receptors, and dominant negative activity. *J Biol Chem*, 271(45): 28235-42.

Yin, J. W., et Wang G. (2014) The Mediator complex: a master coordinator of transcription and cell lineage development. *Development*, 141(5): 977-87.

Yokoyama, A., Katsura, S., Ito, R., Hashiba, W., Sekine, H., Fujiki, R., et Kato, S. (2011) Multiple post-translational modifications in hepatocyte nuclear factor 4 α . *Biochem Biophys Res Commun*, 410(4): 749-53.

Yoon, J. C., Puigserver, P., Chen, G., Donovan, J., Wu, Z., Rhee, J., Adelmant, G., Stafford, J., Kahn, C. R., Granner, D. K., Newgard, C. B., et Spiegelman, B. M. (2001) Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature*, 413(6852): 131-8.

Yuan, X., Ta, T. C., Lin, M., Evans, J. R., Dong, Y., Bolotin, E., Sherman, M. A., Forman, B. M., et Sladek, F. M. (2009) Identification of an endogenous ligand bound to a native orphan nuclear receptor. *PLoS One*, 4(5):e5609.

Zhou, W., Hannoun, Z., Jaffray, E., Medine, C. N., Black, J. R., Greenhough, S., Zhu, L., Ross, J. A., Forbes, S., ... Hay, D. C. (2012) SUMOylation of HNF4 α regulates protein stability and hepatocyte function. *J Cell Sci*, 125(Pt 15): 3630-5.

ANNEXE A - NOMENCLATURE DES RÉCEPTEURS NUCLÉAIRES

Tableau 9 : Nomenclature classique et uniformisée des récepteurs nucléaires, ainsi que leurs ligands identifiés jusqu'à présent (Germain *et al.*, 2006)

Classique	Uniformisée	Ligands
TR α	NR1A1	Hormones thyroïdiennes
TR β	NR1A2	Hormones thyroïdiennes
RAR α	NR1B1	Acide rétinoïque
RAR β	NR1B2	Acide rétinoïque
RAR γ	NR1B3	Acide rétinoïque
PPAR α	NR1C1	Acides gras, leucotriène B ₄ , fibrates
PPAR β	NR1C2	Acides gras
PPAR γ	NR1C3	Acides gras, prostaglandine J ₂ , thiazolidinediones
Rev-erba	NR1D1	Orphelin
Rev-erb β	NR1D2	Orphelin
ROR α	NR1F1	Cholestérol, sulfate de cholestérol
ROR β	NR1F2	Acide rétinoïque
ROR γ	NR1F3	Orphelin
LXR α	NR1H3	Oxystérols, T0901317, GW3965
LXR β	NR1H2	Oxystérols, T0901317, GW3965
FXR α	NR1H4	Acides biliaires, fexaramine
VDR	NR1I1	Vitamine D, 1,25-dihydroxyvitamin D ₃
PXR	NR1I2	Xénobiotiques, 16 α -cyanoprégénolone
CAR	NR1I3	Xénobiotiques, phénobarbital
HNF4 α	NR2A1	Orphelin
HNF4 γ	NR2A2	Orphelin
RXR α	NR2B1	Acide rétinoïque
RXR β	NR2B2	Acide rétinoïque
RXR γ	NR2B3	Acide rétinoïque
TR2	NR2C1	Orphelin
TR4	NR2C2	Orphelin
TLL	NR2E2	Orphelin
PNR	NR2E3	Orphelin
COUP-TFI	NR2F1	Orphelin
COUP-TFII	NR2F2	Orphelin
EAR2	NR2F6	Orphelin
Era	NR3A1	Estradiol-17 β , tamoxifène, raloxifène
Er β	NR3A2	Estradiol-17 β
ERR α	NR3B1	Orphelin
ERR β	NR3B2	DES, 4-OH tamoxifène
ERR γ	NR3B3	DES, 4-OH tamoxifène
GR	NR3C1	Cortisol, dexaméthasone, RU486
MR	NR3C2	Aldostérone, spironolactone
PR	NR3C3	Progestérone, acétate de médroxyprogestérone, RU486
AR	NR3C4	Testostérone, flutamide
NGFI-B	NR4A1	Orphelin
NURR1	NR4A2	Orphelin
NOR1	NR4A3	Orphelin
SF1	NR5A1	Orphelin
LRH-1	NR5A2	Orphelin
GCNF	NR6A1	Orphelin
DAX-1	NR0B1	Orphelin
SHP	NR0B2	Orphelin

ANNEXE B - LA SÉQUENCE RÉELLE DE HNF4a4

En 1996, Drewes et collaborateurs sont les premiers à cloner une isoforme contenant une séquence additionnelle de 90 pb dans le domaine A/B par rapport à l'isoforme HNF4a2 (Figure 23), et qui ne change donc pas le cadre de lecture du reste de la protéine. Les auteurs avancent que cette séquence serait un autre exon du gène *HNF4A*, et serait spécifique à cette isoforme qu'ils dénomment HNF4a4.

HNF4 α 2	ATGCGACTCTCCAAAACCCCTCGTCGACATGGACATGGCCGACTACAGTGCTGCACTGGAC	60
HNF4 α 4	ATGCGACTCTCCAAAACCCCTCGTCGACATGGACATGGCCGACTACAGTGCTGCACTGGAC	60

HNF4 α 2	CCAGCCTACACCACCCTGGAATTTGAGAATGTGCAGGTGTTGACGATGGGCAATG-----	115
HNF4 α 4	CCAGCCTACACCACCCTGGAATTTGAGAATGTGCAGGTGTTGACGATGGGCAATGATTG	120

HNF4 α 2	-----	115
HNF4 α 4	TTGCCGCTGCGTCTCGCCAGATTGAGGCATCCCCCTCCGACATCACTGGAGCATATCTGGA	180
HNF4 α 2	-----ACACGTCCCCATCAGAAGGCACCAACCTCAACGCG	150
HNF4 α 4	GGGGTGGACAGTTCTCCACAGGGAGACACGTCCCCATCAGAAGGCACCAACCTCAACGCG	240

HNF4 α 2	CCCAACAGCCTGGGTGTCAGCGCCCTGTGTGCCATCTGCGGGGACCGGGCCACGGGC...	210
HNF4 α 4	CCCAACAGCCTGGGTGTCAGCGCCCTGTGTGCCATCTGCGGGGACCGGGCCACGGGC...	300

Figure 23 : Alignement de la séquence de l'isoforme HNF4a4 publiée initialement avec la séquence de l'isoforme HNF4a2. Cette séquence montre que la différence entre ces deux isoformes consiste en l'ajout d'une séquence de 90 pb dans le domaine A/B de a4 (Drewes *et al.*, 1996).

Cependant, lorsqu'on compare cette séquence à celle retrouvée dans le chromosome contenant le gène *HNF4A* (20q13.12), on retrouve un nucléotide T supplémentaire au début de l'exon 4. L'ajout d'un nucléotide mène à l'obtention d'un exon de 91 pb, causant un changement du cadre de lecture +1 pour le reste de la protéine, ainsi que l'apparition précoce d'un codon stop (TGA) (Figure 24).

Exon 3

ATGCGACTCTCCAAAACCCCTCGTCGACATGGACATGGCCGACTACAGTGCTGCACTGGACCCAGCCTACACCACCCCT
GGAATTTGAGAAATGTGCAGGTGTTGACGATGGGCAATG

Exon 4

ATTTTGTGTGCCGCTGCGTCTCGCCAGATTGAGGCATCCCCCTCCGACATCACTGGAGCATATCTGGAGGGGTGGACAG
TTCTCCACAGGGAG

Exon 5

ACACGTCCCCATCAGAAGGCACCAACCTCAACGCGCCCAACAGCCTGGGTGTCAGCGCCCTGTGTGCCATCTGCGGG
GACCGGGCCACGGGCAAACACTACGGTGCCCTCGAGCTGTGACGGCTGCAAGGGCTTCTTCCGGAGGAGCGTGCGGAA
GAACCACATGTACTCCTGCAG

HNF4a4 (exon 3 + exon 4 + exon 5) (gène *HNF4A*)

ATGCGACTCTCCAAAACCCCTCGTCGACATGGACATGGCCGACTACAGTGCTGCACTGGACCCAGCCTACACCACCCCT
GGAATTTGAGAAATGTGCAGGTGTTGACGATGGGCAATGATTTTGTGTGCCGCTGCGTCTCGCCAGATTGAGGCATCCC
CTCCGACATCACTGGAGCATATCTGGAGGGGTGGACAGTTCTCCACAGGGAGACACGTCCCCATCAGAAGGCACCAA
CCTCAACGCGCCCAACAGCCTGGGTGTCAGCGCCCTGTGTGCCATCTGCGGGGACCGGGCCACGGGCAAACACTACG
GTGCCTCGAGCTGTGA CGGCTGCAAGGGCTTCTTCCGGAGGAGCGTGCGGAAGAACCACATGTACTCCTGCAG

Figure 24 : Reconstitution de la séquence de HNF4a4 à partir du gène humain *HNF4A*. Le codon initiateur ATG est indiqué en surbrillance verte. Le nucléotide T additionnel est présent au début de la séquence de l'exon 4. L'ajout de ce nucléotide mène à l'apparition précoce d'un codon stop, indiqué en surbrillance rouge.

Il est donc fort probable que dans le papier original datant de 1996, une erreur de séquençage ou d'amplification soit survenue (surtout qu'il s'agit dans ce cas d'une répétition de quatre thymines).

L'isoforme a4 est en réalité traduite à partir d'un codon initiateur alternatif en aval. Cet autre codon ATG n'est pas en phase avec le codon initiateur des isoformes a1/2/3, causant un changement du cadre de lecture -1. Ce codon initiateur n'est donc pas utilisé pour la traduction des autres isoformes, puisqu'il cause l'apparition d'un codon stop de façon prématurée. Un exemple de cette situation pour l'isoforme a2 est présenté ci-dessous (Figure 25).

Exon 3

ATGCGACTCTCCAAAACCCCTCGTCGACATGGACATGGCCGACTACAGTGCTGCACTGGACCCAGCCTACACCACCT
GGAATTTGAGAAATGTGCAGGTGTTGACGATGGGCAATG

Exon 5

ACACGTCCCCATCAGAAGGCACCAACCTCAACGCGCCCAACAGCCTGGGTGTCAGCGCCCTGTGTGCCATCTGCGGG
GACCGGGCCACGGGCAAACACTACGGTGCCTCGAGCTGTGACGGCTGCAAGGGCTTCTTCCGGAGGAGCGTGCGGAA
GAACCACATGTACTCCTGCAG

Exon 6

ATTAGCCGGCAGTGCGTGGTGGACAAAGACAAGAGGAACCAGTGCCGCTACTGCAGGCTCAAGAAATGCTTCCGGG
CTGGCATGAAGAAGGAAG

HNF4 α 2 - ATG alternatif (exon 3 + exon 5 + exon 6)

ATGCGACTCTCCAAAACCCCTCGTCGACATGGACATGGCCGACTACAGTGCTGCACTGGACCCAGCCTACACCACCT
GGAATTTGAGAAATGTGCAGGTGTTGACGATGGGCAATGACACGTCCCCATCAGAAGGCACCAACCTCAACGCGCCCA
ACAGCCTGGGTGTCAGCGCCCTGTGTGCCATCTGCGGGGACCGGGCCACGGGCAAACACTACGGTGCCCTCGAGCTGT
GACGGCTGCAAGGGCTTCTTCCGGAGGAGCGTGCGGAAGAACCACATGTACTCCTGCAGATTAGCCGGCAGTGCGT
GGTGGACAAAGACAAGAGGAACCAGTGCCGCTACTGCAGGCTCAAGAAATGCTTCCGGGCTGGCATGAAGAAGGAAG

Figure 25 : L'utilisation du codon initiateur alternatif pour l'isoforme α 2 produit une protéine tronquée. Le codon initiateur ATG alternatif est indiqué en surbrillance verte. L'utilisation de cet ATG mène à l'apparition d'un codon stop précoce, indiqué en surbrillance rouge.

Toutefois, dans le cas de l'isoforme α 4, l'ajout de l'exon 4 annule ce changement de cadre de lecture (exon 4 +1 ; ATG alternatif -1) (Figure 26A). Ainsi, contrairement à la séquence publiée initialement pour HNF4 α 4, celui-ci possède un domaine A/B entièrement différent par rapport aux isoformes HNF4 α 1/2/3, mais les autres domaines sont identiques entre ces isoformes. L'alignement des séquences protéiques de α 2 et α 4 est présenté dans la figure 26B.

A

Exon 3

ATGCGACTCTCCAAAACCCCTCGTCGACATGGACATGGCCGACTACAGTGCTGCACTGGACCCAGCCTACACCACCT
GGAATTTGAGAATGTGCAGGTGTTGACGATGGGCAATG

Exon 4

ATTTTGTGCGCTGCGTCTCGCCAGATTGAGGCATCCCCTCCGACATCACTGGAGCATATCTGGAGGGGTGGACAG
TTCTCCACAGGGAG

Exon 5

ACACGTCCCCATCAGAAGGCACCAACCTCAACGCGCCCAACAGCCTGGGTGTCAGCGCCCTGTGTGCCATCTGCGGG
GACCGGGCCACGGGCAAACACTACGGTGCCTCGAGCTGTGACGGGTGCAAGGGCTTCTTCCGGAGGAGCGTGCAGAA
GAACCACATGTACTCCTGCAG

HNF4 α 4 - ATG alternatif (exon 3 + exon 4 + exon 5) (gène *HNF4A*)

ATGCGACTCTCCAAAACCCCTCGTCGACATGGACATGGCCGACTACAGTGCTGCACTGGACCCAGCCTACACCACCT
GGAATTTGAGAATGTGCAGGTGTTGACGATGGGCAATGATTTTGTGCGCTGCGTCTCGCCAGATTGAGGCATCCC
CTCCGACATCACTGGAGCATATCTGGAGGGGTGGACAGTTCTCCACAGGGAGACACGTCCCCATCAGAAGGCACCAA
CCTCAACGCGCCCAACAGCCTGGGTGTCAGCGCCCTGTGTGCCATCTGCGGGGACCGGGCCACGGGCAAACACTACG
TGCTCTCGAGCTGTGACGGGTGCAAGGGCTTCTTCCGGAGGAGCGTGCAGGAAGAACACATGTACTCCTGCAG...

B

HNF4 α 2	MRLSKTLVDMADYSAAALDPAYTTLEFENVQVLTMGNDTSPSEGTNLNAPNSLGVSA	60
HNF4 α 4*	----MILLPLRLARLRHPLRH---HWSISGGVDSSPQGDTSPEGTNLNAPNSLGVSA	53
	: : :	
HNF4 α 2	AICGDRATGKHYGASSCDGCKGFFRRSVRKNHMYSCRFSRQCVVDKDKRNQCRYCRL...	120
HNF4 α 4*	AICGDRATGKHYGASSCDGCKGFFRRSVRKNHMYSCRFSRQCVVDKDKRNQCRYCRL...	113

Figure 26 : La séquence réelle de l'isoforme HNF4 α 4. (A) Le codon initiateur ATG alternatif est indiqué en surbrillance verte. L'utilisation de cet ATG permet la conservation du cadre de lecture à travers le reste de la séquence. **(B)** Les isoformes α 2 et α 4 possèdent un domaine A/B complètement distinct, mais les autres domaines sont les mêmes entre ces isoformes.

ANNEXE C - ANNOTATIONS DES ISOFORMES DE HNF4A À TRAVERS DIFFÉRENTES BASES DE DONNÉES

Tableau 10 : Identification des isoformes de HNF4α dans les bases de données RefSeq, Uniprot et Ensembl

Isoforme HNF4α	Séquence (a.a.)	Identifiants et annotations		
		RefSeq	Uniprot	Ensembl
HNF4α1	464	NP_849180.1 Isoform 1 (also known as HNF4alpha1)	P41235-2 HNF4-Alpha-2	ENSP00000412111.1 HNF4A-204
HNF4α2	474	NP_000448.3 Isoform 2 (also known as HNF4alpha2)	P41235-1 HNF4-Alpha-1	ENSP00000312987.3 HNF4A-201
HNF4α3	417	NP_849181.1 Isoform 3 (also known as HNF4alpha3)	P41235-3 HNF4-Alpha-3	ENSP00000410911.2 HNF4A-205
HNF4α4	467	NP_001245284.1 Isoform 7	P41235-4 HNF4-Alpha-4*	-
HNF4α5 (prédite)	457	-	-	-
HNF4α6 (prédite)	420	-	-	-
HNF4α7	442	NP_001025174.1 Isoform 4 (also known as HNF4alpha7)	P41235-6 HNF4-Alpha-8	ENSP00000396216.1 HNF4A-206
HNF4α8	452	NP_787110.2 Isoform 5 (also known as HNF4alpha8)	P41235-5 HNF4-Alpha-7	ENSP00000315180.4 HNF4A-202
HNF4α9	395	NP_001025175.1 Isoform 6 (also known as HNF4alpha9)	P41235-7 HNF4-Alpha-9	ENSP00000476609.1 HNF4A-208
HNF4α10	439	NP_001274111.1 Isoform 8 (also known as HNF4alpha10)	B9VVT8 HNF4alpha10/11/12† <i>Partial sequence</i>	ENSP00000476310.1 HNF4alpha10/11/12† <i>Partial sequence</i>
HNF4α11	449	NP_001274112.1 Isoform 9 (also known as HNF4alpha11)	A0A087WXV4 Hepatocyte nuclear factor 4-alpha‡	ENSP00000481331.1 HNF4A-209
HNF4α12	392	NP_001274113.1 Isoform 10 (also known as HNF4alpha12)	-	-

* Séquence incorrecte de 504 a.a. (ne considère pas le codon initiateur alternatif en aval comparativement à HNF4α1-2-3)

† Séquence partielle de 107 a.a. couvrant le domaine NH2 commun pour HNF4α10-11-12

‡ Séquence complète de HNF4α11, non annotée

ANNEXE D - SÉQUENCES DES PLASMIDES HNF4A

Séquence commune aux 12 isoformes de HNF4a suite à l'optimisation des codons (domaines C, D et E)

GACACTAGTCCAAGCGAGGGGACAAACCTGAACGCACCTAACAGCCTGGGGGTGAGTGCCTTGTGCGCCATTT
GCGGGGACAGAGCAACCGGCAAGCACTACGGAGCCTCAAGCTGTGACGGGTGCAAGGGCTTCTTTCGGCGCAG
CGTGCGGAAGAACCACATGTACAGCTGTGCGTTTAGTTCGGCAGTGCCTGGTCGATAAGGACAAGAGAAACCAG
TGCCGGTACTGCCGCTGAAGAAGTGCTTCCGGGCTGGAATGAAGAAGGAGGCCGTCCAGAACGAACGGGACC
GGATTAGCACCAGGAGAAGCTCTTACGAGGACAGCTCACTCCCCAGCATCAACGCCCTGCTGCAGGCCGAGGT
GCTGTCCAGGCAGATCACATCCCCGTGTGTCGGGCATTAATGGGGACATCAGGGCCAAGAAGATTGCCTCTATT
GCTGACGTGTGCGAGTCCATGAAAGAGCAGCTCCTGGTCCTTGTGGAATGGGCCAAATACATCCAGCCTTTT
GCGAGTTGCCCTGGATGATCAGGTGGCCTTGCTGAGGGCCCATGCCGGGAGCATCTGTTGCTGGGCGCTAC
AAAGAGATCCATGGTCTTTAAGGATGTGTTGCTGCTCGGCAACGACTACATTGTGCCTCGGCACTGCCAGAG
CTTGCCGAGATGAGCCGGGTCTCCATTAGGATTTTGGATGAGCTGGTCCTGCCCTTTCAGGAAGTGCAGATTG
ACGATAACGAGTACGCTTACCTGAAGGCCATCATCTTCTTCGACCCTGACGCCAAGGGCCTGTCCGATCCAGG
GAAAATTAAGCGCCTGAGGAGCCAGGTGCAGGTGTCACTGGAGGACTATATCAATGATCGCCAATACGATTCC
AGAGGGAGGTTCCGGGAGCTCCTCCTCCTGCTGCCCACACTGCAGTCTATCACCTGGCAGATGATCGAGCAGA
TTCAGTTCATTAAGCTCTTTGGCATGGCAAAGATTGACAACCTGTTGCAGGAGATGCTGCTTGGCGGC

Séquence du domaine A/B des isoformes HNF4a1/2/3

ATGCGACTCTCCAAAACCCTCGTAGACATGGACATGGCCGACTACAGTGCTGCACTGGACCCAGCCTACACCA
CCCTGGAATTTGAGAAATGTGCAGGTGTTGACGATGGGCAAT

Séquence du domaine A/B des isoformes HNF4a4/5/6

ATGATTTTGTGCGCTGCGTCTCGCCAGATTGAGGCATCCCCTCCGACATCACTGGAGCATATCTGGAGGGG
TGGACAGTTCTCCACAGGGA

Séquence du domaine A/B des isoformes HNF4a7/8/9

ATGGTCAGCGTGAACGCGCCCCCTCGGGGCTCCAGTGGAGAGTTCTTAC

Séquence du domaine A/B des isoformes HNF4a10/11/12

ATGTCGGAAGTGGGGCCAGGGCTTCCCCCAGGACCCACCA

Séquence du domaine F des isoformes HNF4a1/5/7/10

TCCCCCAGCGATGCACCCCATGCCCACCACCCCCTGCACCCTCACCTGATGCAGGAACATATGGGAACCAACG
 TCATCGTTGCCAACACAATGCCCACTCACCTCAGCAACGGACAGATGTCCACCCCTGAGACCCACAGCCCTC
 ACCGCCAGGTGGCTCAGGGTCTGAGCCCTATAAGCTCCTGCCGGGAGCCGTCGCCACAATCGTCAAGCCCCTC
 TCTGCCATCCCCCAGCCGACCATCACCAAGCAGGAAGTTATC

Séquence du domaine F des isoformes HNF4a2/4/8/11

TCCCCCAGCGATGCACCCCATGCCCACCACCCCCTGCACCCTCACCTGATGCAGGAACATATGGGAACCAACG
 TCATCGTTGCCAACACAATGCCCACTCACCTCAGCAACGGACAGATGTGTGAGTGGCCCCGACCCAGGGGACA
 GGCAGCCACCCCTGAGACCCACAGCCCTCACCGCCAGGTGGCTCAGGGTCTGAGCCCTATAAGCTCCTGCCG
 GGAGCCGTCGCCACAATCGTCAAGCCCCTCTCTGCCATCCCCCAGCCGACCATCACCAAGCAGGAAGTTATC

Séquence du domaine F des isoformes HNF4a3/6/9/12

CCATGTCAAGCACAGGAGGGTCGGGGTTGGAGTGGGGACTCCCCAGGAGACAGGCCTCACACAGTGAGCTCAC
 CCCTCAGCTCCTTGGCTTCCCCACTGTGCCGCTTTGGGCAAGTTGCT

Séquence du plasmide pgLAP5.2-HNF4a1 (GFP)

GACGGATCGGGAGATCTCCCGATCCCCATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAG
 CCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGG
 CAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGG
 CCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAG
 CCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCCGC
 CCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
 AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGT
 CAATGACGGTAAATGGCCCGCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTAC
 ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGT
 TTGACTCACGGGGATTTCGAAGTCTCCACCCCATGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG
 GGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGGTACGGTGGGAGGTC
 TATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGATCGTCGACGAGCTCGTT
 TAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATC
 CAGCCTCCGACTCTAGCGTTTAACTTAAGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACC
 AATTCTCTAAGGAAATACCTTAACCAAGGTCGACATGCGACTCTCCAAAACCTCGTAGACATGGACATGGCCG
 ACTACAGTGCTGCACTGGACCCAGCCTACACCACCTGGAATTTGAGAATGTGCAGGTGTTGACGATGGGCAA
 TGACACTAGTCCAAGCGAGGGGACAAACCTGAACGCACCTAACAGCCTGGGGGTGAGTGCCTTGTGCGCCATT
 TGCGGGGACAGAGCAACCGGCAAGCACTACGGAGCCTCAAGCTGTGACGGGTGCAAGGGCTTCTTTCCGGCGCA
 GCGTGCGGAAGAACCACATGTACAGCTGTGCGTTTAGTCGGCAGTGCGTGGTCGATAAAGGACAAGAGAAACCA
 GTGCCGGTACTGCCGCTGAAGAAGTGCTTCCGGGTGGAATGAAGAAGGAGGCCGTCCAGAACGAACGGGAC
 CGGATTAGCACCAGGAGAAGCTCTTACGAGGACAGCTCACTCCCCAGCATCAACGCCCTGCTGCAGGCCGAGG
 TGCTGTCCAGGCAGATCACATCCCCTGTGTCCGGCATTAAATGGGGACATCAGGGCCAAGAAGATTGCTCTAT
 TGCTGACGTGTGCGAGTCCATGAAAGAGCAGCTCCTGGTCTTGTGGAATGGGCCAAATACATCCCAGCCTTT
 TGCGAGTTGCCCTGGATGATCAGGTGGCCTTGCTGAGGGCCCATGCCGGGAGCATCTGTTGCTGGGCGCTA
 CAAAGAGATCCATTGGTCTTTAAGGATGTGTTGCTGCTCGGCAACGACTACATTGTGCTCGGCACTGCCCAGA
 GCTTGCCGAGATGAGCCGGGTCTCCATTAGGATTTTGGATGAGCTGGTCTGCCCCTTTCAGGAACCTGCAGATT
 GACGATAACGAGTACGCTTACCTGAAGGCCATCATCTTCTTCGACCCCTGACGCCAAGGGCTGTCCGATCCAG
 GGAAAATTAAGCGCTGAGGAGCCAGGTGCAGGTGTCACTGGAGGACTATATCAATGATCGCCAATACGATTTC
 CAGAGGGAGGTTCCGGGAGCTCCTCCTCTGCTGCCACACTGCAGTCTATCACCTGGCAGATGATCGAGCAG
 ATTCAAGTTCATTAAGCTCTTTGGCATGGCAAAGATTGACAACCTGTTGCAGGAGATGCTGCTTGGCGGCTCCC
 CCAGCGATGCACCCCATGCCACCACCCCTGCACCCTCACCTGATGCAGGAACATATGGGAACCAACGTCAT
 CGTTGCCAACACAATGCCCACTCACCTCAGCAACGGACAGATGTCCACCCCTGAGACCCACAGCCCTCACCG
 CCAGGTGGCTCAGGGTCTGAGCCCTATAAGCTCCTGCCGGGAGCCGTGCCACAATCGTCAAGCCCTCTCTG
 CCATCCCCCAGCCGACCATCACCAAGCAGGAAGTTATCCACTCGAGATATCTAGACCCAGCTTTCTTGTAACA
 AGTGGTTGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGCGGTTCCGAAAAAGAAACCGCTGCT
 GCTAAATTCGAACGCCAGCACATGGACAGCAGCAGCGGCCCTGGAAGTTCTGTTCCAGGGGCCCTGGAAGTTC
 GTTCCAGGGGCCCTCCGGAGGGATGGTGATGGTGAGCAAGGGCGAGGAGCTGTTACACCGGGGTGGTGCCCAT
 CCTGGTTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTACAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACC
 TACGGCAAGCTGACCCCTGAAGTTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCCACCCCTCGTGACCA
 CCCTGACCTACGGCGTGCACTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGC
 CATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG
 GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACA
 TCCTGGGGCACAACTGAGTACAACCTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAGCGG
 CATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAG
 AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCCTACCTGAGCACCCAGTCCGCCCCGAGCA
 AAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTTCGTGACCGCCGCGGGGATCACTCTCGGCAT
 GGACGAGCTGTACAAGTAAATTCGAAGGTAAGCCTATCCCTAACCCCTCTCCTCGGTCTCGATTCTACGCGTACC
 GGTTAGTAATGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCTTCTAGTTGCCAGCCATCTGTTGTTTGGC
 CCTCCCCCGTGCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTCTCCTAATAAAATGAGGAAATTCG
 ATCGCATTGTCTGAGTAGGTGTCTATTCTATTTCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG
 GAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCT
 CTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGAC
 CGTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCTTTCTCGCCACGTTTCGCCGGC
 TTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCA

AAAAATTGATTAGGGTGATGGTTCACGTACCTAGAAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAAGTAT
 AGGAACTTCCTTGGCCAAAAAGCCTGAACTCACC GCGACGTCTGTGAGAAAGTTTCTGATCGAAAAAGTTCGAC
 AGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTACAGCTTCGATGTAGGAGGGCGTG
 GATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACCTTTGCATC
 GGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTACAGCGAGAGCCTGACCTATTGCATCTCCCGC
 CGTGACACAGGGTGTACGTTGCAAGACCTGCCGAAACCGAACTGCCCCGTGTTCTGCAGCCGGTCGCGGAGG
 CCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAAGGAATCGG
 TCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACTGTGATG
 GACGACACCGTCAGTGCGTCCGTGCGCGAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAG
 TCCGGCACCTCGTGACGCGGATTTCCGGCTCCAACAATGTCTGACGGACAATGGCCGCATAACAGCGGTTCAT
 TGACTGGAGCGAGGCGATGTTCCGGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGAGGCGCGTGTTG
 GCTTGATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCGAGGATCGCCGCGGCTCCGGG
 CGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTTCGATGATGCGAGCTTG
 GGCGCAGGGTTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTGCGGCGTACACAAATCGCCCGCAGA
 AGCGCGGCCGCTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAACCGACGCCCCAGCACTCGTC
 CGAGGGCAAAGGAATAGCACGTACTACGAGATTTTCGATTCCACCGCCGCCCTTCTATGAAAGGTTGGGCTTCGG
 AATCGTTTTCCGGGACGCGCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCCACCCC
 AACTTGTTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATT
 TTTCACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTC
 TAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCCTGTGTGAAATTGTTATCCGCTCACAATCCACAC
 AACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGT
 TGCGCTCACTGCCCCGCTTTCAGTTCGGGAAACCTGTCTGTCCAGCTGCATTAATGAATCGGCCAACGCGCGGG
 GAGAGGCGGTTTGCCTATTGGGCGCTCTTCCGCTTCCCTCGCTCACTGACTCGCTGCGCTCGGTCTCGGCTG
 CGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGA
 ACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCT
 CCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGA
 TACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGT
 CCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCCGTTGTAGGT
 CGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCCCAGCGCTGCGCCTTATCCGGTAACTAT
 CGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAG
 CGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACCTACGGCTACACTAGAAAGAACAGTATT
 TGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACC
 ACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATC
 CTTTGATCTTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATT
 ATCAAAAAGGATCTTCACCTAGATCCTTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATATGAG
 TAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCTGTTTCAT
 CCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGCCCCAGTGCTGC
 AATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAG
 CGCAGAAGTGGTCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTA
 GTTCGCCAGTTAATAGTTTGCACAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGG
 TATGGCTTCATTACGCTCCGGTTCCTCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCG
 GTTAGCTCCTTCGGTCTCCGATCGTTGTGAGAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAG
 CACTGCATAATTCTCTTACTGTGATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTC
 ATTCTGAGAATAGTGATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAAATACCGCGCCACAT
 AGCAGAACTTTAAAAAGTGCTCATCATTGGAAAAAGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGT
 TGAGATCCAGTTCGATGTAACCCACTCGTGACCCAACTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTC
 TGGGTGAGCAAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTC
 ATACTCTTCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAAT
 GTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAAGTGCCACCTGACGTC

Séquence du plasmide pgLAP5.2-HNF4a1-BioID2-3Xmyc (BioID2) (3Xmyc)

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACCTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAG
CCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGG
CAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGG
CCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAG
CCCATATATGGAGTTCCGCGTTACATAAATTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCCCGC
CCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGT
CAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTAC
ATCTACGTATTAGTCATCGCTATTACCATGGTGTATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGT
TTGACTCACGGGGATTTCCAAAGTCTCCACCCCAATTGACGTCAATGGGAGTTTTGTTTTTGGCACCAAAATCAACG
GGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTC
TATATAAGCAGAGCTCTCCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGATCGTCGACGAGCTCGTT
TAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATC
CAGCCTCCGGACTCTAGCGTTTAAACTTAAGGCAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATAGAACC
AATTCTCTAAGGAAATACTTAACCAAGGTCGACATGCGACTCTCCAAAACCCTCGTAGACATGGACATGGCCG
ACTACAGTGTGCTGCACTGGACCCAGCCTACACCACCCTGGAATTTGAGAATGTGCAGGTGTTGACGATGGGCAA
TGACACTAGTCCAAGCGAGGGGACAAACCTGAACGCACCTAACAGCCTGGGGGTGAGTGCCTTGTGCGCCATT
TGCGGGGACAGAGCAACCGCAAGCACTACGGAGCCTCAAGCTGTGACGGGTGCAAGGGCTTCTTTCGCGCGCA
GCGTGCGGAAGAACCACATGTACAGCTGTGCGTTTAGTTCGGCAGTGCGTGGTCGATAAGGACAAGAGAAACCA
GTGCCGGTACTGCCGCTGAAGAAGTGCTTCCGGGCTGGAATGAAGAAGGAGGCCGTCCAGAACGAACGGGAC
CGGATTAGCACCAGGAGAAGCTCTTACGAGGACAGCTCACTCCCCAGCATCAACGCCCCTGCTGCAGGCCGAGG
TGCTGTCCAGGCAGATCACATCCCCCTGTGTCCGGCATTAAATGGGGACATCAGGGCCAAGAGATTGCCCTCTAT
TGCTGACGTGTGCGAGTCCATGAAAGAGCAGCTCCTGGTCCTTGTGGAATGGGCCAAATACATCCCAGCCTTT
TGCGAGTTGCCCCTGGATGATCAGGTGGCCTTGTCTGAGGGCCATGCCGGGAGCATCTGTTGTCTGGGCGCTA
CAAGAGATCCATGGTCTTTAAGGATGTGTTGCTGCTCGCAACGACTACATTGTGCTCGGCACTGCCAGA
GCTTGCCGAGATGAGCCGGGTCTCCATTAGGATTTTGGATGAGCTGGTCTGCCCTTTTCAGGAACTGCAGATT
GACGATAACGAGTACGCTTACCTGAAGGCCATCATCTTCTTCGACCCTGACGCCAAGGGCCTGTCCGATCCAG
GGAAAATTAAGCGCCTGAGGAGCCAGGTGCAAGTGTCACTGGAGGACTATATCAATGATCGCCAATACGATTTC
CAGAGGGAGGTTTCGGGGAGCTCCTCCTCTGCTGCCACACTGCAGTCTATCACCTGGCAGATGATCGAGCAG
ATTCAAGTTTCAATTAAGCTCTTTTGGCATGGCAAAGATTGACAACCTGTTGCAGGAGATGCTGCTTGGCGGCTCCC
CCAGCGATGCACCCCATGCCACCACCCCTGCACCCTCACCTGATGCAGGAACATATGGGAACCAACGTCAT
CGTTGCCAACACAATGCCCACTCACCTCAGCAACGGACAGATGTCCACCCCTGAGACCCACAGCCCTCACCG
CCAGGTGGCTCAGGGTCTGAGCCCTATAAGCTCCTGCCGGGAGCCGTCGCCACAATCGTCAAGCCCTCTCTG
CCATCCCCCAGCCGACCATACCAAGCAGGAAGTTATCCACTCGAGATATCTAGACCCAGCTTTCTTGTACAA
AGTGGTTGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGGGCCCGCGGTTGAAAAAGAAACCGCTGCT
GCTAAATTGCAACGCCAGCACATGGACAGCAGCAGCGCCTGGAAGTTCTGTTCCAGGGGCCCTGGAAGTTC
TGTTCCAGGGGCCCTCCGGAGGGATGGTGGGTACCTTCAAGAACCCTGATCTGGCTGAAGGAGGTGGACAGCAC
CCAGGAGAGACTGAAGGAGTGGAACTGAGCTACGGCACCAGCCCTGGTGGCCGACAGACAGACCAAGGGCAGA
GGCGGCCCTGGGCAGAAAGTGGCTGAGCCAGGAGGGCGGCCGTACTTCAGCTTCTGCTAAACCCGAAGGAGT
TCGAAAAATCTACTGCAGCTTCCCCCTGGTGTGTTGGGCTTAGCGTGAGTGAAGCCCTGGAGGAGATCACCGAGAT
CCCCCTCAGCCTGAAGTGGCCCAACGACGTGTACTTCCAGGAGAAGAAGGTGAGCGGCGTGTGTGCGAGCTG
AGCAAGGACAAGCTGATCGTGGGCATCGGCATCAACGTGAACCAGAGAGAGATCCCCGAGGAGATCAAGGACA
GAGCCACCACCCTGTACGAGATCACCGGCAAGGACTGGGACAGAAAAGGAGGTGCTGCTGAAGGTGCTGAAGAG
AATCAGCGAGAACCCTGAAGAAGTTCAAGGAGAAGAGCTTCAAGGAGTTCAAGGGCAAGATCGAGAGCAAGATG
CTGTACCTGGGCGAGGAGGTGAAGCTGTGGGCGAGGGCAAGATCACCGGCAAGCTGGTGGGCTGAGCGAGA
AGGGCGGCGCCCTGATCCTGACCGAGGAGGGCATCAAGGAGATCCTGAGCGGCGAGTTGAGCCTGAGAAGAAG
CGACGAGCAGAACTCATCTCTGAAGAAGATCTGGAACAAAAGTTGATTTTCAGAAGAAGATCTGGAACAGAAG
CTCATCTCTGAGGAAGATCTGTGAAGCTTTTCGAAGGTAAAGCCTATCCCTAACCCCTCCTCGGTCTCGATT
CTACGCGTACCGGTTAGTAATGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCCTTCTAGTTGCCAGCCATC
TGTTGTTTTGCCCCCTCCCCCGTGCCTTCCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCCTTTCTTAATAAAAT
GAGGAAATTCATCGCATTGTCTGAGTAGGTGTCTATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG
GGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAAC
CAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACG
CGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTCTTCCCTTCTTCTCGCCA

CGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCA
 CCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTACCTAGAAAGTTCCTATTCGGAAGTTCCTATTCT
 CTAGAAAGTATAGGAACTTCCTTGGCCAAAAAGCCTGAACTCACCGCGACGTCTGTCGAGAAAGTTTCTGATCG
 AAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGT
 AGGAGGGCGTGGATATGTCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGG
 CACTTTGCATCGGCCGCGCTCCCATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATT
 GCATCTCCCGCCGTGCACAGGGTGTACGTTGCAAGACCTGCCGAAAACCGAACTGCCCGCTGTTCTGCAGCC
 GGTCGCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCCGGCCATTCCGGACCG
 CAAGGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGC
 AAATGTGATGGACGACACCGTCAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGA
 CTGCCCCGAAGTCCGGCACCTCGTGACGCGGATTTCCGGCTCCAACAATGTCTGACGGACAATGGCCGCATA
 ACAGCGGTCAATTGACTGGAGCGAGGCGATGTTCCGGGATTCCCAATACGAGGTGCCAACATCTTCTTCTGGA
 GGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCGCC
 GCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGTTGACGGCAATTTTCGAT
 GATGCAGCTTGCGCGCAGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAA
 TCGCCCCGAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAAGTACTCGCCGATAGTGGAACCGACGCC
 CAGCACTCGTCCGAGGGCAAAGGAATAGCACGTACTACGAGATTTTCGATTCCACCGCCGCCCTTCTATGAAAGG
 TTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCT
 TCGCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAAATAAGCAATAGCATCACAAATTTACAAA
 TAAAGCATTTTTTCTACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGTCTGTATA
 CCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTTGTTATCCGCTCA
 CAATTCACACAACATAACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCAC
 ATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAAATGAATCGGC
 CAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCCTCGCTCACTGACTCGCTGCGCTCGG
 TCGTTCCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAATACGGTTATCCACAGAATCAGGGGATAA
 CGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTT
 TTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAG
 GACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCTGTTCCGACCCTGCCGCTTAC
 CGGATACCTGTCCGCTTTTCTCCCTTCGGGAAGCGTGGCGCTTTTCTCATAGCTCAGCTGTAGGTATCTCAGT
 TCGGTGTAGGTTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCGTTTCAGCCCGACCGCTGCGCCTTAT
 CCGGTAACTATCGTCTTGAGTCCAACCCGTTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAACAG
 GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGA
 AGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCG
 GCAAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTTCGAAGCAGCAGATTACGCGCAGAAAAAAGGATC
 TCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTG
 GTCATGAGATTATCAAAAAGGATCTTACCTAGATCCTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAA
 GTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCT
 ATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTGCTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGC
 CCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCAGCTCCAGATTTATCAGCAATAAACCAGCCAGCCG
 GAAGGGCCGAGCGCAGAAAGTGGTCTGCAACTTTATCCGCCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGC
 TAGAGTAAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCG
 TCGTCTGTTTGGTATGGCTTCATTTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGT
 GCAAAAAAGCGTTAGCTCCTTCCGGTCTCCGATCGTTGTGAGAAGTAAGTTGGCCGCGAGTTTATCACTCAT
 GGTATGGCAGCACTGCATAATTCTCTTACTGTGATGCCATCCGTAAAGATGCTTTTCTGTGACTGGTGAGTAC
 TCAACCAAGTCATTCTGAGAATAGTGATGCGGCGACCGAGTTGCTCTTGCCCGCGCTCAATACGGGATAATA
 CCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTTGGAACAGTTCTTCCGGGCGAAAACTCTCAAGGAT
 CTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGACCCCACTGATCTTCAGCATCTTTTACTTTT
 ACCAGCTTTTCTGGGTGAGCAAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAT
 GTTGAATACTCATACTCTTCTTTTCAATAATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATA
 CATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAC
 GTC

ANNEXE E - CONDITIONS PCR

Tableau 11 : Oligonucléotides utilisés pour l'amplification de la séquence commune de HNF4a

Oligonucléotide	Séquence (5' → 3')
Séquence commune HNF4a - sens	TCGCTGGTCGACTGCGCTACTAGTCCAAGCGAGGGGACAAACCTGAAC
Séquence commune HNF4a - antisens	ACTTGTCTCGAGTCATTGGGCGCCAAGCAGCATCTCCTGCAACAGG

Tableau 12 : Réactifs utilisés pour l'amplification de la séquence commune de HNF4a

Réactif	Volume (µl)
ADN matrice (pLenti-HNF4a2-GFP) - 10 ng/µl	1,0
Oligonucléotide sens - 10 µM (IDT, San Jose, États-Unis)	1,25
Oligonucléotide antisens - 10 µM (IDT, San Jose, États-Unis)	1,25
dNTPs - 10 mM (Wisent, Saint-Jean-Baptiste, Canada)	0,5
Tampon HF - 5X (Bio-Rad, Hercules, États-Unis)	5,0
H ₂ O stérile	15,75
ADN polymérase iProof High-Fidelity (Bio-Rad, Hercules, États-Unis)	0,25

Tableau 13 : Paramètres PCR d'amplification de la séquence commune de HNF4a

Étape	Nombre de cycles	Température (°C)	Durée (s)
Dénaturation initiale		98	180
Hybridation initiale	2	61	30
Élongation initiale		72	30
Dénaturation		98	10
Hybridation	28	72	20
Élongation		72	30
Élongation finale	1	72	300

Tableau 14 : Oligonucléotides utilisés pour l'amplification du plasmide pgLAP5.2

Oligonucléotide	Séquence (5' → 3')
Sens	GGGCCCTCCGGAGGGATGGTG
Antisens	GAGAGGGTTAGGGATAGGCTTACCTTCGAA

Tableau 15 : Oligonucléotides utilisés pour amplifier le vecteur pgLAP5.2-BioID2-3Xmyc-vide

Oligonucléotide	Séquence (5' → 3')
Sens	ATGTTCAAGAACCTGATCTGGCTG
Antisens	/Phosphorylation 5' /CGGTGTCTTCTATGGAGGTC

Tableau 16 : Paramètres PCR d'amplification du vecteur pgLAP5.2-BioID2-3Xmyc-vide

Étape	Nombre de cycles	Température (°C)	Durée (s)
Dénaturation initiale	1	98	180
Dénaturation		98	10
Hybridation	30	58	20
Élongation		72	180
Élongation finale	1	72	300

Tableau 17 : Liste des oligonucléotides utilisés pour l'amplification des gènes partenaires
MS

Nom des amorces (gène)	Direction	Séquence (5' → 3')	Taille du produit (pb)
<i>BRD8</i>	Sens	GGGGACAAGTTTGTACAAAAAGCAGGCTTCACCATGGCGACGGGAACGGGC	2917
	Antisens	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTTTTCATCTTCATATCTGCTTCAATGGC	
<i>EPC1</i>	Sens	GGGGACAAGTTTGTACAAAAAGCAGGCTTCACCATGAGTAAACTGTCGTTTCGG	2572
	Antisens	GGGGACCACTTTGTACAAGAAAGCTGGGTCCGTCACCTCCATCGCTAC	
<i>FOSL1</i>	Sens	GGGGACAAGTTTGTACAAAAAGCAGGCTTCACCATGTTCCGAGACTTCGGG	877
	Antisens	GGGGACCACTTTGTACAAGAAAGCTGGGTCCAAAGCGAGGAGGGTTGG	
<i>GATAD2B</i>	Sens	GGGGACAAGTTTGTACAAAAAGCAGGCTTCACCATGGATAGAATGACAGAAGATGCTC	1843
	Antisens	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTTCTGTCCACTGATGGACTG	
<i>HNF4G</i>	Sens	GGGGACAAGTTTGTACAAAAAGCAGGCTTCACCATGGACATGGCAAATTACAGTG	1399
	Antisens	GGGGACCACTTTGTACAAGAAAGCTGGGTCCAATTGCTTTTGTGTTGGAGAGATGC	
<i>ING3</i>	Sens	GGGGACAAGTTTGTACAAAAAGCAGGCTTCACCATGTTGTACCTAGAAGACTATCTGG	1318
	Antisens	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTTGTGTCTGCTGCCCTCTTC	
<i>IRF2BP2</i>	Sens	GGGGACAAGTTTGTACAAAAAGCAGGCTTCACCATGGCCGCGCGGTGGCG	1825
	Antisens	GGGGACCACTTTGTACAAGAAAGCTGGGTCCGAGTCTCTCTCTTTTTCACTTTCACATC	
<i>JUN</i>	Sens	GGGGACAAGTTTGTACAAAAAGCAGGCTTCACCATGACTGCAAAGATGGAAACG	1057
	Antisens	GGGGACCACTTTGTACAAGAAAGCTGGGTCAAATGTTTGCAACTGCTGCG	
<i>MTA1</i>	Sens	GGGGACAAGTTTGTACAAAAAGCAGGCTTCACCATGGCCGCCAACATGTAC	2209
	Antisens	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGTCCTCGATGACGATGGG	
<i>SMARCA4</i>	Sens	GGGGACAAGTTTGTACAAAAAGCAGGCTTCACCATGTCCACTCCAGACCCAC	5101
	Antisens	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGCTTTCTTCGCTGCCACTTC	
<i>SMARCD2</i>	Sens	GGGGACAAGTTTGTACAAAAAGCAGGCTTCACCATGTGGGCGGAGGCGCG	1657
	Antisens	GGGGACCACTTTGTACAAGAAAGCTGGGTCTGCTCAGGCGAATTCAGCACCTG	
<i>ZNF629</i>	Sens	GGGGACAAGTTTGTACAAAAAGCAGGCTTCACCATGGAGCCCAGACTGCG	2671
	Antisens	GGGGACCACTTTGTACAAGAAAGCTGGGTCCAGGGAGACACCTGGGTG	

Tableau 18 : Réactifs utilisés pour l'amplification des gènes partenaires MS

iProof High-Fidelity		TransStart KD Plus	
Réactif (iProof)	Volume (µl)	Réactif (iProof)	Volume (µl)
ADNc (dilution 1/20)	5,0	ADNc (dilution 1/20)	5,0
Oligonucléotide sens - 10 µM (IDT, San Jose, États-Unis)	1,0	Oligonucléotide sens - 10 µM (IDT, San Jose, États-Unis)	1,13
Oligonucléotide antisens - 10 µM (IDT, San Jose, États-Unis)	1,0	Oligonucléotide antisens - 10 µM (IDT, San Jose, États-Unis)	1,13
dNTPs - 10 mM (Wisent, Saint-Jean-Baptiste, Canada)	0,4	dNTPs - 2.5 mM (TransGen Biotech, Beijing, Chine)	1,8
Tampon GC - 5X (Bio-Rad, Hercules, États-Unis)	4,0	Tampon KD - 5X (TransGen Biotech, Beijing, Chine)	4,0
DMSO	0,6	H ₂ O stérile	6,5
H ₂ O stérile	11,8	ADN polymérase TransStart KD Plus (TransGen Biotech, Beijing, Chine)	0,45
ADN polymérase iProof (Bio-Rad, Hercules, États-Unis)	0,2		

Tableau 19 : Paramètres d'amplification des gènes partenaires MS

Nom du gène	Enzyme ADN polymérase	Dénaturation initiale	Dénaturation	Hybridation	Élongation	Élongation finale
		1 cycle	35 cycles			1 cycle
		Température (°C) et durée ('minutes, ''secondes)				
BRD8	iProof High-Fidelity	98 (3')	98 (10'')	65 (20'')	72 (1'10'')	72 (5')
EPC1	iProof High-Fidelity	98 (3')	98 (10'')	65 (20'')	72 (1'10'')	72 (5')
FOSL1	iProof High-Fidelity	98 (3')	98 (10'')	68 (20'')	72 (30'')	72 (5')
GATAD2B	iProof High-Fidelity	98 (3')	98 (10'')	64 (20'')	72 (50'')	72 (5')
HNF4G	TransStart KD Plus	94 (3')	94 (30'')	54 (30'')	68 (1'30'')	68 (5')
ING3	iProof High-Fidelity	98 (3')	98 (10'')	64 (20'')	72 (35'')	72 (5')
IRF2BP2	iProof High-Fidelity	98 (3')	98 (10'')	70 (20'')	72 (50'')	72 (5')
JUN	iProof High-Fidelity	98 (3')	98 (10'')	68 (20'')	72 (30'')	72 (5')
MTA1	iProof High-Fidelity	98 (3')	98 (10'')	70 (20'')	72 (1')	72 (5')
SMARCA4	TransStart KD Plus	94 (3')	94 (30'')	60 (30'')	68 (5')	68 (5')
SMARCD2	iProof High-Fidelity	98 (3')	98 (10'')	70 (20'')	72 (50'')	72 (5')
ZNF629	iProof High-Fidelity	98 (3')	98 (10'')	70 (20'')	72 (1'10'')	72 (5')

Tableau 20 : Liste des oligonucléotides utilisés pour l'amplification des 12 isoformes de HNF4a et des gènes de référence

Nom des amorces (gène)	Direction	Séquence (5' → 3')	Taille du produit (pb)
<i>HNF4A-1</i>	Sens	CAGCCTACACCACCCTGGAATTTG	1206
	Antisens	GGGTCTCAGGGGTGGACATC	
<i>HNF4A-2</i>	Sens	CAGCCTACACCACCCTGGAATTTG	1206
	Antisens	GTCGGGGCCACTCACACATC	
<i>HNF4A-3</i>	Sens	CAGCCTACACCACCCTGGAATTTG	1169
	Antisens	CAGTGGGGAAGCCAAGGAGC	
<i>HNF4A-4</i>	Sens	CCGACATCACTGGAGCATATCTGGAG	1202
	Antisens	GTCGGGGCCACTCACACATC	
<i>HNF4A-5</i>	Sens	CCGACATCACTGGAGCATATCTGGAG	1202
	Antisens	GGGTCTCAGGGGTGGACATC	
<i>HNF4A-6</i>	Sens	CCGACATCACTGGAGCATATCTGGAG	1165
	Antisens	CAGTGGGGAAGCCAAGGAGC	
<i>HNF4A-7</i>	Sens	GGGCTCCAGTGGAGAGTTCTTACG	1176
	Antisens	GGGTCTCAGGGGTGGACATC	
<i>HNF4A-8</i>	Sens	GGGCTCCAGTGGAGAGTTCTTACG	1176
	Antisens	GTCGGGGCCACTCACACATC	
<i>HNF4A-9</i>	Sens	GGGCTCCAGTGGAGAGTTCTTACG	1139
	Antisens	CAGTGGGGAAGCCAAGGAGC	
<i>HNF4A-10</i>	Sens	ATGTCGGACTGGGGCCAG	1192
	Antisens	GGGTCTCAGGGGTGGACATC	
<i>HNF4A-11</i>	Sens	ATGTCGGACTGGGGCCAG	1192
	Antisens	GTCGGGGCCACTCACACATC	
<i>HNF4A-12</i>	Sens	ATGTCGGACTGGGGCCAG	1155
	Antisens	CAGTGGGGAAGCCAAGGAGC	
<i>HPRT</i>	Sens	ACCAGTCAACAGGGGACATAA	169
	Antisens	CTTCGTGGGGTCCTTTTCACC	
<i>PUM1</i>	Sens	TGAGGTGTGCACCATGAAC	187
	Antisens	CAGAATGTGCTTGCCATAGG	
<i>POLR2A</i>	Sens	GATTGTGAATGGGGATGACC	1257
	Antisens	ATAAGCACGTCCACCGTTTC	
<i>PSMB2</i>	Sens	GCCCCGACTATGTTCTTGTC	1144
	Antisens	GTCAGGATGGTCTCGCTCTC	

Tableau 21 : Réactifs utilisés pour l'amplification des isoformes de HNF4a par PCR

Réactif	Volume (µl)
ADNc (dilution 1/20)	2,0
ADN polymérase Taq (Service de purification de protéines, Université de Sherbrooke)	0,2
Tampon Taq - 10X	2,0
dNTPs - 25 mM (Amersham Biosciences, Little Chalfont, Royaume-Uni)	0,16
Oligonucléotide sens - 10 µM (IDT, San Jose, États-Unis)	1,0
Oligonucléotide antisens - 10 µM (IDT, San Jose, États-Unis)	1,0
H ₂ O stérile	13,64

Tableau 22 : Paramètres PCR d'amplification des isoformes de HNF4a

Étape	Nombre de cycles	Température (°C)	Durée (s)
Dénaturation initiale	1	94	240
Dénaturation		94	30
Hybridation	32	62	30
Élongation		72	75
Élongation finale	1	72	300

Tableau 23 : Liste des oligonucléotides utilisés pour l'amplification par RT-qPCR

Nom des amorces (gène)	Direction	Séquence (5' → 3')	Taille du produit (pb)
<i>HNF1A</i>	Sens	GCACAGCTTGGAGCAGACAT	192
	Antisens	GTTGATGACCGGCACACTCT	
<i>VIL1</i>	Sens	AGCCAGATCACTGCTGAGGT	163
	Antisens	AGGTGTTTCCTCCTTCCTGCT	
<i>APOA1</i>	Sens	TGGATGTGCTCAAAGACAGC	216
	Antisens	CCAGATCCTTGCTCATCTCC	
<i>CREB3L3</i>	Sens	CCTGGAGACTCGGATGTCAG	186
	Antisens	GCAAAGGACAGCAACAGGAC	
<i>HNF4A</i>	Sens	ACCTGAACGCACCTAACAGC	198
	Antisens	GGCACTGGTTTCTCTTGTC	
<i>TBP</i>	Sens	GGGGAGCTGTGATGTGAAGT	130
	Antisens	GGAGAACAATTCTGGGTTTGA	

Tableau 24 : Réactifs utilisés pour l'amplification par RT-qPCR

Réactif	Volume (µl)
ADNc (dilution 1/20)	2,0
Fast Start SYBR Green Master (Roche Applied Science, Penzberg, Allemagne)	10
Oligonucléotide sens - 10 µM (IDT, San Jose, États-Unis)	1,0
Oligonucléotide antisens - 10 µM (IDT, San Jose, États-Unis)	1,0
H ₂ O stérile	6,0

Tableau 25 : Conditions d'amplification par RT-qPCR

Étape	Nombre de cycles	Température (°C)	Durée (' minutes, "secondes)
Préincubation	1	95	10'
Dénaturation	45	95	10"
Hybridation		60	10"
Élongation		72	20"
		95	10"
Courbe de fusion	1	65	60"
		97	1"
Refroidissement	1	37	30"

ANNEXE F - LECTURES DE SÉQUENÇAGE DE L'ARN

Tableau 26 : Nombre de lectures obtenues par séquençage Novaseq pour chaque échantillon exprimant ou non les isoformes de HNF4a en triplicatas

Isoforme	Échantillon	Nombre de lectures (pb)	Isoforme	Échantillon	Nombre de lectures (pb)
Vide	A0N1	56 055 654	$\alpha 7$	A7N1	47 161 043
	A0N2	79 014 778		A7N2	53 814 942
	A0N3	51 144 357		A7N3	51 384 496
$\alpha 1$	A1N1	59 796 558	$\alpha 8$	A8N1	56 834 615
	A1N2	61 102 956		A8N2	45 719 495
	A1N3	50 640 711		A8N3	50 602 340
$\alpha 2$	A2N1	56 296 640	$\alpha 9$	A9N1	48 612 904
	A2N2	59 046 798		A9N2	66 235 540
	A2N3	55 091 536		A9N3	46 139 778
$\alpha 3$	A3N1	55 233 772	$\alpha 10$	A10N1	49 359 919
	A3N2	54 825 664		A10N2	63 943 754
	A3N3	58 118 727		A10N3	58 612 012
$\alpha 4$	A4N1	64 325 507	$\alpha 11$	A11N1	49 679 058
	A4N2	60 119 152		A11N2	56 464 567
	A4N3	56 840 108		A11N3	54 190 854
$\alpha 5$	A5N1	52 032 011	$\alpha 12$	A12N1	61 130 462
	A5N2	60 313 773		A12N2	57 330 031
	A5N3	54 432 846		A12N3	50 603 261
$\alpha 6$	A6N1	48 678 898			
	A6N2	47 471 097			
	A6N3	50 688 423			

ANNEXE G - ENRICHISSEMENT SILAC DES SOUS-UNITÉS DES COMPLEXES BAF/PBAF

Tableau 27 : Enrichissement par protéomique des membres des complexes de remodelage de la chromatine SWI/SNF BAF/PBAF pour les 12 isoformes de HNF4a

Membres BAF/PBAF	Ratios d'enrichissement normalisés											
	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 4$	$\alpha 5$	$\alpha 6$	$\alpha 7$	$\alpha 8$	$\alpha 9$	$\alpha 10$	$\alpha 11$	$\alpha 12$
<i>ARID1A</i>	4,93	4,48	6,43	6,35	5,28	4,13	4,09	4,19	4,67	6,46	6,57	8,33
<i>ARID1B</i>	5,09	3,90	6,73	10,39	8,76	7,63	3,81	3,97	4,26	4,91	5,52	7,03
<i>ARID2</i>	4,42	3,57	5,59	5,18	5,71	4,38	4,27	3,72	4,54	5,71	5,25	6,48
<i>BCL11B</i>	3,04	2,23	3,31	7,68	6,83	7,62	2,63	3,03	2,96	3,17	3,53	4,05
<i>BRD7</i>	3,92	3,50	4,10	6,49	4,83	5,06	3,41	3,27	3,97	2,82	2,99	3,94
<i>BRD9</i>	5,66	5,37	4,97	8,65	-	7,77	1,72	1,54	1,56	2,66	1,78	2,37
<i>DPF2</i>	4,65	3,67	6,35	5,39	4,36	3,75	4,64	3,83	5,43	6,93	6,58	9,62
<i>GLTSCR1</i>	2,86	2,61	-	3,52	2,98	-	2,89	1,77	-	2,88	3,39	-
<i>GLTSCR1L</i>	2,85	2,86	-	3,32	2,30	-	2,78	2,63	-	3,42	4,05	-
<i>PBRM1</i>	3,05	3,31	3,53	3,12	3,15	4,23	2,87	2,92	3,61	4,63	4,98	5,45
<i>PHF10</i>	-	2,01	-	-	-	1,57	-	3,56	-	-	-	3,77
<i>SMARCA4</i>	4,16	3,96	4,54	4,57	3,92	3,00	4,04	4,12	3,82	6,00	6,69	6,80
<i>SMARCC1</i>	3,64	3,35	4,68	3,92	3,97	3,18	3,53	3,81	3,98	5,25	5,43	6,81
<i>SMARCC2</i>	2,69	4,10	3,69	2,95	3,32	2,85	2,37	3,02	3,29	3,40	3,81	5,21
<i>SMARCD2</i>	4,32	3,85	4,92	4,42	6,15	3,78	4,85	4,65	5,00	7,16	8,87	8,57
<i>SMARCE1</i>	2,99	2,45	4,30	2,94	2,85	2,74	2,97	2,69	3,41	3,87	3,84	4,02